

GenCore version 5.1.6
Copyright (C) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2005, 16:58:19 ; Search time 23.6993 Seconds
(without alignments)
375.348 Million cell updates/sec

Title: US-09-998-861-7

Perfect score: 102

Sequence: 1 XTKTFFSVKDIILSLPEQXKXCA 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A: Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	97	95.1	23	5	ABP53054
2	68	66.7	318	7	AD537113
3	68	66.7	318	7	AD537113
4	68	66.7	319	5	AAO21474
5	68	66.7	324	4	AAG64848
6	68	66.7	324	4	AAG64318
7	68	66.7	324	4	AAB99919
8	68	66.7	324	7	AD537113
9	68	66.7	324	7	AD537113
10	61	59.8	364	7	ADG76983
11	60	58.8	23	5	ABP53048
12	60	58.8	86	2	AAV25174
13	60	58.8	273	2	AAV25173
14	60	58.8	273	4	AAW78480
15	60	58.8	273	5	ABP70157
16	60	58.8	273	8	ADQ09849
17	60	58.8	279	4	AAW79464
18	59	57.8	134	4	ABG14069
19	55	53.9	416	4	AB863565
20	54	52.9	281	5	AAG77977
21	54	52.9	301	5	AAG77976
22	50	49.0	344	4	ABO65415
23	47	46.1	786	8	ADN26956
24	46.5	45.6	371	2	AAH83015
25	46.5	45.6	371	4	AAU05583

26	46.5	45.6	401	7	AD537113	Thyroid t
27	46.5	45.6	401	8	AD537113	Human mal
28	46	45.1	23	5	ABP53050	GIP relat
29	46	45.1	365	8	ADQ09852	Mouse NK-
30	45	44.1	171	4	ABG19376	Novel hum
31	45	44.1	321	6	ABM67666	Phototrab
32	45	44.1	338	7	ADF06776	Bacterial
33	44	43.1	197	4	ABG04855	Novel hum
34	44	43.1	322	8	ADN99762	Novel hum
35	44	43.1	349	5	ABO9760	Amino aci
36	44	43.1	382	4	AB863575	Drosophil
37	44	43.1	471	4	ABG04859	Novel hum
38	44	43.1	553	2	AAW03528	Salmonell
39	44	43.1	732	8	ADN21330	Bacterial
40	43	42.2	87	8	ADQ09850	Mouse NK-
41	43	42.2	731	7	ABO73778	Pseudomon
42	43	42.2	1006	4	ABR71507	Drosophil
43	43	42.2	1819	5	AAW47588	Drosophil
44	42.5	41.7	94	3	AAH14154	Arabidops
45	42.5	41.7	254	7	ABO66470	Klebsiell

ALIGNMENTS

RESULT 1

ABP53054

ID ABP53054 standard; peptide; 23 AA.

XX AC ABP53054;

XX AC ABP53054;

DT 06-NOV-2002 (first entry)

XX GIP related peptide Nkx consensus SEQ ID NO:7.

XX Human; Groucho-interacting protein; GIP; Nkx; Nkx6.3; differentiation;

XX Groucho corepressor protein.

OS Rattus sp.

OS Gallus sp.

OS Mus sp.

OS Synthetic.

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

CC A purified complex (PC1), comprising a first polypeptide comprising a GIP
 CC and the second polypeptide comprising a Groucho corepressor protein, can
 CC be used for identifying an agent which modulates the stability or
 CC activity of the complex, or that disrupts a polypeptide complex. PC1 is
 CC useful for screening the candidate substance interacting with the complex
 CC or with (1); for inhibiting the guided differentiation of a cell
 CC resulting in the impairment of ventral patterning; for identifying a
 CC polypeptide complex in a subject; and determining altered expression of a
 CC polypeptide in a subject. M1 is useful for guiding the fate of
 CC differentiation of a cell into a specific cell type. The present sequence
 CC represents a GIP related peptide, which is used in the exemplification of
 CC the present invention.

XX SQ Sequence 23 AA;

Query Match 95.1%; Score 97; DB 5; Length 23;
 Best Local Similarity 100.0%; Pred. No. 9.9e-10;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TKTPFSVKDILSLPQXRXAGA 23
 DB 2 TKTPFSVKDILSLPQXRXAGA 23

RESULT 2
 ADE54613
 ID ADE54613 standard; protein; 318 AA.

XX AC ADE54613;

XX DT 29-JAN-2004 (first entry)

XX DE Rat Protein O35767, SEQ ID NO 418.

XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

XX OS Rattus norvegicus.

XX WO2003016475-A2.

XX 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.

XX 14-AUG-2001; 2001US-0312147P.

XX 01-NOV-2001; 2001US-0346382P.

XX 26-NOV-2001; 2001US-0333347P.

XX (GEO) GEN HOSPITAL CORP.

XX (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-268312/26.

XX GENBANK; O35767.

XX New composition comprising two or more isolated polypeptides, useful for
 XX preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
 XX or human polynucleotides or a polynucleotide which represents a fragment,
 XX derivative or allelic variation of the nucleic acid sequence. Also
 XX claimed are a vector comprising the novel polynucleotide, a host cell
 XX comprising the vector, a method for identifying a nucleotide sequence
 XX which is differentially regulated in an animal subjected to pain and a
 XX kit to perform the method, an array, a method for identifying an agent
 XX that increases or decreases the expression of the polynucleotide sequence
 XX that is differentially expressed in neuronal tissue of a first animal
 XX subjected to pain, a method for identifying a compound which regularly
 XX the expression of a polynucleotide sequence which is differentially

CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene
 CC injury (CCI) and spared nerve injury (SNI)).
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 318 AA;

Query Match 66.7%; Score 68; DB 7; Length 318;
 Best Local Similarity 63.6%; Pred. No. 0.0027;
 Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 TKTPFSVKDILSLPQXRXAGA 23
 DB 8 TKTPFSVKDILSLPQXRXAGA 29

RESULT 3

ADE54617

ID ADE54617 standard; protein; 318 AA.

XX AC ADE54617;

XX DT 29-JAN-2004 (first entry)

XX DE Rat Protein O35767, SEQ ID NO 422.

XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

XX OS Rattus norvegicus.

XX WO2003016475-A2.

XX 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.

XX 14-AUG-2001; 2001US-0312147P.

XX 01-NOV-2001; 2001US-0346382P.

XX 26-NOV-2001; 2001US-0333347P.

XX (GEO) GEN HOSPITAL CORP.

XX (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-268312/26.

XX GENBANK; O35767.

XX New composition comprising two or more isolated polypeptides, useful for
 XX preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
 XX or human polynucleotides or a polynucleotide which represents a fragment,
 XX derivative or allelic variation of the nucleic acid sequence. Also
 XX claimed are a vector comprising the novel polynucleotide, a host cell
 XX comprising the vector, a method for identifying a nucleotide sequence
 XX which is differentially regulated in an animal subjected to pain and a
 XX kit to perform the method, an array, a method for identifying an agent


```

AAG64318
ID AAG64318 standard; protein; 324 AA.
XX
AC AAG64318;
XX
DT 24-SEP-2001 (first entry)
XX
DE Human Nkx2.5/Csx protein.
XX
KW Angiogenesis; cardiant; cell differentiating agent; bone marrow;
KW heart muscle cell; heart disease; human; Nkx2.5/Csx.
XX
OS Homo sapiens.
XX
PN WO200148149-A1.
XX
PD 05-JUL-2001.
XX
PF 28-FEB-2000; 2000WO-JP001148.
XX
PR 28-DEC-1999; 99JP-00372826.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Umezawa A, Hata J, Fukuda K, Ogawa S, Sakurada K;
DR WPI; 2001-418252/44.
DR N-PSDB; AAH49590.
XX
PS New adult bone marrow-originated cells capable of differentiating into
XX heart muscle cells, applicable as remedies for various heart diseases
XX particularly with damaged heart muscle accompanying degeneration.
XX
XX Claim 22; Page 70-71; 158pp; Japanese.
XX
CC The present invention relates to cells isolated from bone marrow, which
CC are capable of at least differentiating into heart muscle cells. The
CC cells are applicable as remedies for various heart diseases particularly
CC with damaged heart muscle accompanying degeneration. The present sequence
CC was used to illustrate the present invention
XX
SQ Sequence 324 AA;
Query Match 66.7%; Score 68; DB 4; Length 324;
Best Local Similarity 63.6%; Pred. No. 0.0028;
Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 2 TKTPTFSVKDILSLPEQXRAXGA 23
DB 8 TPTPTFSVKDILNLQEQQRSLAA 29
RESULT 7
AAB99919
ID AAB99919 standard; protein; 324 AA.
XX
AC AAB99919;
XX
DT 26-SEP-2001 (first entry)
XX
DE Human Nkx2.5/Csx protein sequence SEQ ID NO:9.
XX
KW Differentiation; heart muscle cell; cytokine; transcription factor;
KW proliferation; surface antigen; heart disease; cardiomyocyte;
KW bone marrow; umbilical blood cell; heart muscle degeneration;
KW myocardial infarction.
XX
OS Homo sapiens.
XX
PN WO200148150-A1.
XX
PD 05-JUL-2001.
XX
PF 02-NOV-2000; 2000WO-JP007741.
XX
PR 28-DEC-1999; 99JP-00372826.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Umezawa A, Hata J, Fukuda K, Ogawa S, Sakurada K, Gojo S;
DR WPI; 2001-425655/45.
DR N-PSDB; AAH44355.
XX
PS Cells capable of differentiating into cardiomyocytes and originating in
XX bone marrow or umbilical blood cells for study of cardiomyocyte
XX differentiation and treatment of heart disease.
XX
XX Claim 40; Page 95-96; 187pp; Japanese.
XX
CC The present invention describes cells originating in bone marrow or
XX umbilical blood cells which are capable of differentiating into
XX cardiomyocytes. Also described are: (1) cardiomyocytes produced by the
XX differentiation of the cells; (2) a method for carrying out the
XX differentiation into cardiomyocytes, regulated by a promotonal and/or
XX inhibitory factor; (3) a method for the differentiation of the cells into
XX cell types other than cardiomyocytes; (4) drug compositions promoting the
XX formation of heart muscle and regeneration of heart tissue which contain
XX the cells; (5) a method for the production of antibodies which recognise
XX the cells, especially antibodies which recognise a surface antigen on the
XX cells; (6) a method for screening factors which promote the proliferation
XX of the cells; (7) a method for immortalising the cells by expressing
XX telomerase in them; (8) drug compositions for the treatment of heart
XX disease which contain the immortalised cells; and (9) cell-free
XX supernatant from the culture of the cells and its use in promoting their
XX differentiation into cardiomyocytes. The cells are used in the treatment
XX of diseases involving heart muscle degeneration, such as myocardial
XX infarction and in the study of cardiomyocyte differentiation. AAH44351 to
XX AAH44409 and AAB99915 to AAB99935 represent sequences used in the
XX exemplification of the present invention
XX
SQ Sequence 324 AA;
Query Match 66.7%; Score 68; DB 4; Length 324;
Best Local Similarity 63.6%; Pred. No. 0.0028;
Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 2 TKTPTFSVKDILSLPEQXRAXGA 23
DB 8 TPTPTFSVKDILNLQEQQRSLAA 29
RESULT 8
ADE54619
ID ADE54619 standard; protein; 324 AA.
XX
AC ADE54619;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human Protein P52952, SEQ ID NO 424.
XX
KW Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
OS Homo sapiens.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX

```



```

PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
XX (GEO ) GEN HOSPITAL CORP.
XX (FARB ) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
XX GENBANK; P52952.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (SNI)), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 324 AA;
XX
XX Query Match 66.7%; Score 68; DB 7; Length 324;
XX Best Local Similarity 63.6%; Pred. No. 0.0028;
XX Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
XX
QY 2 TKTPTSVDILSLPEQXRAXGA 23
DB 8 TPTPTSVDILNLEQQRSGLAA 29
XX
RESULT 9
ADE54615
ID ADE54615 standard; protein; 324 AA.
XX
XX ADE54615;
XX
XX 29-JAN-2004 (first entry)
XX
XX Human Protein P52952, SEQ ID NO 420.
XX
XX Human; pain; neuronal tissue; gene therapy;
XX spinal segmental nerve injury; chronic constriction injury; CCI;
XX spared nerve injury; SNI; Chung.
XX
XX Homo sapiens.
XX
XX WO2003016475-A2.
XX

```

```

XX
XX 27-FEB-2003.
XX
XX 14-AUG-2002; 2002WO-US025765.
XX
XX 14-AUG-2001; 2001US-0312147P.
XX 01-NOV-2001; 2001US-0346382P.
XX 26-NOV-2001; 2001US-0333347P.
XX
XX (GEO ) GEN HOSPITAL CORP.
XX (FARB ) BAYER AG.
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
XX GENBANK; P52952.
XX
XX New composition comprising two or more isolated polypeptides, useful for
XX preparing a medicament for treating pain in an animal.
XX
XX Claim 1; Page; 1017pp; English.
XX
XX The invention discloses a composition comprising two or more isolated rat
XX or human polynucleotides or a polynucleotide which represents a fragment,
XX derivative or allelic variation of the nucleic acid sequence. Also
XX claimed are a vector comprising the novel polynucleotide, a host cell
XX comprising the vector, a method for identifying a nucleotide sequence
XX which is differentially regulated in an animal subjected to pain and a
XX kit to perform the method, an array, a method for identifying an agent
XX that increases or decreases the expression of the polynucleotide sequence
XX that is differentially expressed in neuronal tissue of a first animal
XX subjected to pain, a method for identifying a compound which regulates
XX the expression of a polynucleotide sequence which is differentially
XX expressed in an animal subjected to pain, a method for identifying a
XX compound that regulates the activity of one or more of the
XX polynucleotides, a method for producing a pharmaceutical composition, a
XX method for identifying a compound or small molecule that regulates the
XX activity in an animal of one or more of the polypeptides given in the
XX specification, a method for identifying a compound useful in treating
XX pain and a pharmaceutical composition comprising the one or more
XX polypeptides or their antibodies. The polynucleotide or the compound that
XX modulates its activity is useful for preparing a medicament for treating
XX pain (e.g. spinal segmental nerve injury (SNI)), chronic constriction
XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
XX therapy). The sequence presented is a human protein (shown in Table 2 of
XX the specification) which is differentially expressed during pain. Note:
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic form directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 324 AA;
XX
XX Query Match 66.7%; Score 68; DB 7; Length 324;
XX Best Local Similarity 63.6%; Pred. No. 0.0028;
XX Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
XX
QY 2 TKTPTSVDILSLPEQXRAXGA 23
DB 8 TPTPTSVDILNLEQQRSGLAA 29
XX
RESULT 10
ADG76983
ID ADG76983 standard; protein; 364 AA.
XX
XX ADG76983;
XX
XX 11-MAR-2004 (first entry)
XX
XX Human nucleic acid associated polypeptide (NAAP) 11.
XX
XX nucleic acid associated polypeptide; NAAP; cytostatic;
XX antiarteriosclerotic; anticonvulsant; nootropic; neuroprotective;
XX

```


PD 04-FEB-1999.
 XX 16-JUL-1998; 98WO-US014765.
 XX 25-JUL-1997; 97US-00900510.
 XX 20-JAN-1998; 98US-00008892.
 XX (RECC) UNIV CALIFORNIA.
 PA (ARCH-) ARCH DEV CORP.
 XX German MS, Rubenstein JLR, Sussel L, Sander M, Hartigan-O'Connor DJ, Pedersen RA, Meneses JJ, Bell GI, Furuta H;
 XX WPI; 1999-142918/12.
 DR N-PSDB; AAY25174.
 XX Novel human Nkx-2.2 DNA and protein - useful for generating transgenic animals as models and for diagnosing predisposition to e.g. diabetes, depression or obesity.
 XX Claim 2; Fig 1A-B; 58pp; English.
 CC This invention describes a novel human Nkx-2.2 polypeptide and its encoding nucleic acid. Nkx-2.2 is a member of a vertebrate gene family homologous to the Drosophila NK2 gene, which provides initial positional information for specific ventral regions in the developing forebrain. The gene is also expressed in the pancreas, pancreatic islet-beta cells and hamster insulinoma cells. The polynucleotides can be used to identify homologous polynucleotides using probes derived from the polynucleotide sequence. Transgenic animals produced from the products of the invention can be used to screen for antagonists. Analysis of an individual for a Nkx-2.2 polymorphism can be used to diagnose a predisposition to an Nkx-2.2-associated disorder, especially diabetes, depression and obesity.
 CC This sequence represents a fragment of the human Nkx2.2 protein corresponding to exon 1 of the encoding nucleotide sequence
 XX Sequence 86 AA;
 SQ Query Match 58.8%; Score 60; DB 2; Length 86;
 Best Local Similarity 59.1%; Pred. No. 0.016;
 Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 2 TKTGFSVKDILSLPQXRAXGA 23
 DB 6 TKTGFSVKDILDLPTNDEGS 27
 RESULT 13
 AAY25173
 ID AAY25173 standard; protein; 273 AA.
 XX AAY25173;
 AC AAY25173;
 DT 10-SEP-1999 (first entry)
 XX Human Nkx2.2 protein.
 DE Nkx2.2; human; vertebrate gene family; NK2 homologue; transgenic;
 KW positional information; ventral; developing forebrain; pancreas;
 KW pancreatic islet-beta cell; insulinoma cell; polymorphism; diagnosis;
 KW predisposition; diabetes; depression; obesity.
 XX Homo sapiens.
 OS Homo sapiens.
 XX WO9905258-A1.
 PN WO9905258-A1.
 XX 04-FEB-1999.
 PD 16-JUL-1998; 98WO-US014765.
 XX 25-JUL-1997; 97US-00900510.
 XX 20-JAN-1998; 98US-00008892.

PA (RECC) UNIV CALIFORNIA.
 PA (ARCH-) ARCH DEV CORP.
 XX German MS, Rubenstein JLR, Sussel L, Sander M, Hartigan-O'Connor DJ, Pedersen RA, Meneses JJ, Bell GI, Furuta H;
 XX WPI; 1999-142918/12.
 DR N-PSDB; AAX78505, AAX78523.
 XX Novel human Nkx-2.2 DNA and protein - useful for generating transgenic animals as models and for diagnosing predisposition to e.g. diabetes, depression or obesity.
 XX Claim 2; Fig 1A-B; 58pp; English.
 CC This invention describes a novel human Nkx-2.2 polypeptide and its encoding nucleic acid. Nkx-2.2 is a member of a vertebrate gene family homologous to the Drosophila NK2 gene, which provides initial positional information for specific ventral regions in the developing forebrain. The gene is also expressed in the pancreas, pancreatic islet-beta cells and hamster insulinoma cells. The polynucleotides can be used to identify homologous polynucleotides using probes derived from the polynucleotide sequence. Transgenic animals produced from the products of the invention can be used to screen for antagonists. Analysis of an individual for a Nkx-2.2 polymorphism can be used to diagnose a predisposition to an Nkx-2.2-associated disorder, especially diabetes, depression and obesity.
 CC This sequence represents the human Nkx2.2 protein described in the invention
 XX Sequence 273 AA;
 SQ Query Match 58.8%; Score 60; DB 2; Length 273;
 Best Local Similarity 59.1%; Pred. No. 0.059;
 Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 2 TKTGFSVKDILSLPQXRAXGA 23
 DB 6 TKTGFSVKDILDLPTNDEGS 27
 RESULT 14
 AAM78480
 ID AAM78480 standard; protein; 273 AA.
 XX AAM78480;
 AC AAM78480;
 DT 06-NOV-2001 (first entry)
 XX Human protein SEQ ID NO 1142.
 DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX Homo sapiens.
 OS Homo sapiens.
 XX WO200157190-A2.
 PN WO200157190-A2.
 XX 09-AUG-2001.
 PD 05-FEB-2001; 2001WO-US004098.
 XX 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00663561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX (HYSE-) HYSEQ INC.
 PA

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 DR N-PSDB; AAK51613.
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX Claim 20; Page 3375-3375; 6221pp; English.
 PS The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX
 SQ Sequence 273 AA;

Query Match 58.8%; Score 60; DB 4; Length 273;
 ID AAK51613 standard; protein; 273 AA.
 Best Local Similarity 59.1%; Pred. No. 0.059;
 Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 2 TKTGFSVKDILSLPEQXRAXGA 23
 DB ||| ||||| ||| : : :
 6 TKTGFSVKDILSLPEQXRAXGA 27

RESULT 15
 ABP70157
 ID ABP70157 standard; protein; 273 AA.
 XX
 AC ABP70157;
 XX
 DT 07-APR-2003 (first entry)
 XX
 DE Amino acid sequence of oestrogen receptor alpha cofactor CF18.
 XX
 KW Human; oestrogen receptor alpha; ER-alpha; CF16; CF17; CF18; CF19; CF40;
 KW CF41; CF42; CF43; cofactor; osteoporosis; bone disease; reproduction;
 KW hormonal dysfunction; cancer; cardiovascular disease; atherosclerosis;
 KW hot flush; mood change; Alzheimer's disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200270699-A2.
 XX
 PD 12-SEP-2002.
 XX
 PF 28-FEB-2002; 2002WO-EP002189.
 XX
 PR 01-MAR-2001; 2001EP-00105062.
 XX
 PA (LION-) LION BIOSCIENCE AG.
 XX
 PI Albers M, Ellwanger S, Loeser E, Koegl M;
 XX
 DR WPI; 2002-713451/77.
 DR N-PSDB; AB223376.
 XX

PT New cofactors of estrogen receptor alpha, designated as CF16, CF17, CF18,
 PT CF19, CF40, CF41, CF42 and/or CF43, useful for screening of compounds for
 PT treating osteoporosis, hormonal dysfunctions, cancer or cardiovascular

PT diseases.
 XX Claim 12; Page 72-73; 111pp; English.
 XX The present sequence represents a cofactor of oestrogen receptor alpha
 CC (ER-alpha), designated CF18. The specification describes CF16, CF17,
 CC CF18, CF19, CF40, CF41, CF42, and CF43. The cofactor polypeptides and
 CC nucleic acid molecules are useful for screening for compounds for
 CC treating osteoporosis and other bone diseases, failures in reproductive
 CC functions or hormonal dysfunctions, cancer or cardiovascular diseases
 CC such as atherosclerosis, and in preventing hot flushes, mood changes and
 CC Alzheimer's disease. The CF proteins are also useful for screening for
 CC ligands of the ER alpha. The nucleic acid sequences are useful for making
 CC vectors and CF polypeptides, transforming host cells, as research tools
 CC for developing nucleic acid probes, and for developing analytical tools
 CC such as antisense oligonucleotides
 XX
 SQ Sequence 273 AA;
 Query Match 58.8%; Score 60; DB 5; Length 273;
 Best Local Similarity 59.1%; Pred. No. 0.059;
 Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 2 TKTGFSVKDILSLPEQXRAXGA 23
 DB ||| ||||| ||| : : :
 6 TKTGFSVKDILSLPEQXRAXGA 27

RESULT 16
 ADQ09849
 ID ADQ09849 standard; protein; 273 AA.
 XX
 AC ADQ09849;
 XX
 DT 23-SEP-2004 (first entry)
 XX
 DE Human NK-2 class homeobox protein, NKx2.2.
 XX
 KW Human; islet cell differentiation transcription factor;
 KW insulin-dependent diabetes; insulin; somatic cell;
 KW insulin-producing cell; islet cell; NeuroD; ngn3; neurogenin 3; Pax6;
 KW paired-box transcription factor 6; Pax4;
 KW NKx6.1; Isl-1; Islet factor 1; Pdx-1;
 KW pancreatic and duodenal homeobox protein; BCT; betacellulin.
 XX
 OS Homo sapiens.
 XX
 PN US2004132679-A1.
 XX
 PD 08-JUL-2004.
 XX
 PF 03-SEP-2003; 2003US-00654102.
 XX
 PR 03-SEP-2002; 2002US-0407743P.
 XX
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 XX
 PI Chan L, Kojima H;
 XX
 DR WPI; 2004-517032/49.
 DR N-PSDB; ADQ09866, ADQ09867, ADQ09940, ADQ09941.
 XX
 PT Use of an islet cell differentiation transcription factor polypeptide or
 PT its homologue or analog for treating a mammal for insulin-dependent
 PT diabetes, increasing an insulin level in a somatic cell, or generating an
 PT insulin-producing cell.
 XX
 PS Claim 109; SEQ ID NO 87; 190pp; English.
 XX
 CC The invention relates to the use of an islet cell differentiation
 CC transcription factor polypeptide or its homologue or analogue for
 CC treating a mammal for insulin-dependent diabetes, increasing an insulin

CC level in a somatic cell, or generating an insulin-producing cell. Also
 CC included are a method of treating a mammal for insulin-dependent
 CC diabetes, a method of increasing an insulin level in a somatic cell, a
 CC method of generating an insulin-producing cell, a therapeutic composition
 CC comprising an isolated islet cell differentiation transcription factor
 CC polypeptide (and/or an isolated nucleic acid expressing the polypeptide),
 CC an insulin-producing cell comprising a vector (where the vector comprises
 CC a nucleic acid sequence encoding an islet cell differentiation
 CC transcription factor), an insulin-producing cell (generated by a method
 CC comprising obtaining a somatic cell and transfecting the cell with a
 CC vector comprising a nucleic acid sequence encoding an islet cell
 CC differentiation transcription factor, where in the transfecting step the
 CC cell produces insulin), a method of generating at least one pancreatic
 CC islet, and a composition (comprising: NeuroD or ngn3 polypeptide or a
 CC polynucleotide expressing a NeuroD or ngn3 polypeptide and betacellulin
 CC polypeptide or a polynucleotide expressing a betacellulin polypeptide).
 CC The islet cell differentiation transcription factor polypeptide is
 CC NeuroD, ngn3 (neurogenin 3), Pax6 (paired-box transcription factor 6),
 CC Pax4 (paired-box transcription factor 4), Nkx2.2 (NK-2 class homeobox
 CC protein), Nkx6.1, Isl-1 (islet factor 1), Pdx-1 (pancreatic and duodenal
 CC homeobox protein), Bcr (betacellulin) or their combinations. The islet
 CC cell differentiation transcription factor polypeptide or its homologue or
 CC analogue is useful for treating a mammal for insulin-dependent diabetes,
 CC increasing an insulin level in a somatic cell, or generating an insulin-
 CC producing cell. The present sequence is an islet cell differentiation
 CC transcription factor polypeptide as detailed above.

XX SQ Sequence 273 AA;

Query Match 58.8%; Score 60; DB 8; Length 273;
 Best Local Similarity 59.1%; Pred. No. 0.059; 7; Indels 0; Gaps 0;
 Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 TKTGFSVKDILSLPEQXRAXGA 23
 ||| ||||| ||| :
 Db 6 TKTGFSVKDILDLPTNDEGS 27

RESULT 17

AAAM79464
 ID AAM79464 standard; protein; 279 AA.

AC AAM79464;

XX 06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 3110.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

XX WO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US004098.

XX 03-FEB-2000; 2000US-00496914.

PR 27-APR-2000; 2000US-00560875.

PR 20-JUN-2000; 2000US-00598075.

PR 19-JUL-2000; 2000US-00620325.

PR 01-SEP-2000; 2000US-00654936.

PR 15-SEP-2000; 2000US-00663561.

PR 20-OCT-2000; 2000US-00693325.

PR 30-NOV-2000; 2000US-00728422.

PA (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

PI Ma Y, Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX WPI; 2001-476283/51.
 DR N-PSDB; AAK52597.

XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.

XX Claim 20; Page 254; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication

XX Sequence 279 AA;

Query Match 58.8%; Score 60; DB 4; Length 279;
 Best Local Similarity 59.1%; Pred. No. 0.061;
 Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 TKTGFSVKDILSLPEQXRAXGA 23
 ||| ||||| ||| :
 Db 12 TKTGFSVKDILDLPTNDEGS 33

RESULT 18

ABG14069
 ID ABG14069 standard; protein; 134 AA.

XX AC ABG14069;

XX 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #14060.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639162/73.

DR N-PSDB; AAK52597.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX Claim 20; SEQ ID NO 44428; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 134 AA;
SQ

CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX
XX Sequence 416 AA;
XX
XX Query Match 53.9%; Score 55; DB 4; Length 416;
XX Best Local Similarity 64.7%; Pred. No. 0.74;
XX Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
XX

QY 4 TPFSVKDILSLPEQXRA 20
| | | | | : : |
DB 36 TPFSVKDILMNVNQTEA 52

RESULT 20
AAG77977
ID AAG77977 standard; protein; 281 AA.
XX AAG77977;
XX
XX 18-FEB-2002 (first entry)
XX
XX Human NK-2 homeobox transcription factor Nkx2-C4 #2.
XX
XX Human; NK-2 homeobox transcription factor; Nkx2-C4; cardiac; arrhythmia;
XX antiarrhythmic; hypotensive; acute cardiac failure; cardiotoxicity;
XX chronic cardiac failure; myocardial infarction; cardiac hypertrophy;
XX myocarditis; pulmonary hypertension; coronary heart disease.
XX
XX Homo sapiens.
XX
XX W0200181400-A1.
XX
XX 01-NOV-2001.
XX
XX 11-APR-2001; 2001WO-EP004145.
XX
XX 19-APR-2000; 2000EP-00108485.
XX (MERE) MERCK PATENT GMBH.
XX
XX Wilm C;
XX
XX WPI; 2002-034432/04.
XX N-PSDB; AAK98240.
XX
XX New polypeptide, for diagnosing, treating myocardial infarction,
XX arrhythmia, myocarditis and for identifying modulators of therapeutic
XX use, comprises the novel human NK-2 homeobox transcription factor
XX polypeptides and polynucleotides.
XX
XX Claim 1; Page 37-38; 41pp; English.
XX
XX The sequence represents a novel human NK-2 homeobox transcription factor
XX of the invention. The invention relates to novel isolated human NK-2
XX homeobox transcription factor (Nkx2-C4) polypeptides. The polypeptides of
XX the invention have cardiac, antiarrhythmic, and hypotensive activity and
XX can act as a vaccine or modulator of Nkx2-C4 polypeptide. Nkx2-C4
XX polypeptide is useful in screening assays to identify compounds that
XX stimulate or inhibit the function or level of the polypeptide. Agonists
XX and antagonists of the polynucleotide are useful for treating acute and
XX chronic cardiac failure of different etiologies, myocardial infarction,
XX cardiac hypertrophy, arrhythmia, myocarditis, pulmonary hypertension,
XX cardiotoxicity (e.g. induced by chemotherapy) and coronary heart disease.
XX The polypeptides are useful as immunogens to produce antibodies which are
XX useful for treating diseases, to isolate or to identify clones expressing
XX the polypeptide or to purify the polypeptides by affinity chromatography.

CC Nkx2-C4 polypeptides are also useful to identify membrane bound or
 CC soluble receptors
 XX
 SQ Sequence 281 AA;

Query Match 52.9%; Score 54; DB 5; Length 281;
 Best Local Similarity 68.8%; Pred. No. 0.71;
 Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 TKTPTSVDKILSLPEQ 17
 | | | | | | | | | | :
 Db 7 TSTPTSVDKILSLERE 22

RESULT 21
 AAG77976
 ID AAG77976 standard; protein; 301 AA.
 XX
 AC AAG77976;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Human NK-2 homeobox transcription factor Nkx2-C4 #1.
 XX
 KW Human; NK-2 homeobox transcription factor; Nkx2-C4; cardiant; arrhythmia;
 KW antiarrhythmic; hypotensive; acute cardiac failure; cardiotoxicity;
 KW chronic cardiac failure; myocardial infarction; cardiac hypertrophy;
 KW myocarditis; pulmonary hypertension; coronary heart disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200181400-A1.
 XX
 PD 01-NOV-2001.
 XX
 PF 11-APR-2001; 2001WO-EP004145.
 XX
 PR 19-APR-2000; 2000EP-00108485.
 XX
 PA (MERE) MERCK PATENT GMBH.
 XX
 PI Wilm C;
 XX
 PS WPI; 2002-034432/04.
 DR N-PSDB; AAK98239.
 XX
 CC New polypeptide, for diagnosing, treating myocardial infarction,
 PT arrhythmia, myocarditis and for identifying modulators of therapeutic
 PT use, comprises the novel human NK-2 homeobox transcription factor
 PT polypeptides and polynucleotides.
 XX
 PS Claim 1; Page 34-35; 41pp; English.
 XX
 CC The sequence represents a novel human NK-2 homeobox transcription factor
 CC of the invention. The invention relates to novel isolated human NK-2
 CC homeobox transcription factor (Nkx2-C4) polypeptides. The polypeptides of
 CC the invention have cardiant, antiarrhythmic, and hypotensive activity and
 CC can act as a vaccine or modulator of Nkx2-C4 polypeptide. Nkx2-C4
 CC polypeptide is useful in screening assays to identify compounds that
 CC stimulate or inhibit the function or level of the polypeptide. Agonists
 CC and antagonists of the polynucleotide are useful for treating acute and
 CC chronic cardiac failure of different etiologies, myocardial infarction,
 CC cardiac hypertrophy, arrhythmia, myocarditis, pulmonary hypertension,
 CC cardiotoxicity (e.g. induced by chemotherapy) and coronary heart disease.
 CC The polypeptides are useful as immunogens to produce antibodies which are
 CC useful for treating diseases, to isolate or to identify clones expressing
 CC the polypeptide or to purify the polypeptides by affinity chromatography.
 CC Nkx2-C4 polypeptides are also useful to identify membrane bound or
 CC soluble receptors
 XX
 SQ Sequence 301 AA;

Query Match 52.9%; Score 54; DB 5; Length 301;

Best Local Similarity 68.8%; Pred. No. 0.77;
 Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 TKTPTSVDKILSLPEQ 17
 | | | | | | | | | | :
 Db 7 TSTPTSVDKILSLERE 22

RESULT 22
 ABO65415
 ID ABO65415 standard; protein; 344 AA.
 XX
 AC ABO65415;
 XX
 DT 29-JUL-2004 (first entry)
 XX
 DE Klebsiella pneumoniae polypeptide seqid 11932.
 XX
 KW Recombinant expression vector; transcription regulatory element;
 KW Klebsiella pneumoniae protein; antibacterial; Vaccine.
 XX
 OS Klebsiella pneumoniae.
 XX
 PN US6610836-B1.
 XX
 PD 26-AUG-2003.
 XX
 PF 27-JAN-2000; 2000US-00489039.
 XX
 PR 29-JAN-1999; 99US-0117747P.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Breton GL, Osborne M;
 XX
 DR WPI; 2003-895346/82.
 DR N-PSDB; ACH98966.
 XX
 CC New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
 PT preparing a vaccine composition against Klebsiella pneumoniae.
 XX
 PS Disclosure; SEQ ID NO 11932; 932pp; English.
 XX
 CC The invention describes a new isolated nucleic acid encoding a Klebsiella
 CC pneumoniae polypeptide. Also described are: a recombinant expression
 CC vector comprising the nucleic acid, operably linked to a transcription
 CC regulatory element; and a cell comprising the recombinant expression
 CC vector. The nucleic acid is useful for preparing a vaccine composition
 CC against Klebsiella pneumoniae. This is the amino acid sequence of a
 CC Klebsiella pneumoniae polypeptide of the invention
 XX
 SQ Sequence 344 AA;

Query Match 49.0%; Score 50; DB 7; Length 344;
 Best Local Similarity 57.9%; Pred. No. 4.6;
 Matches 11; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 TPFSVKOILSLPEQXRXG 22
 | | | | | | | | | | :
 Db 87 TPFSVEITLSIEETARANG 105

RESULT 23
 ADN26956
 ID ADN26956 standard; protein; 786 AA.
 XX
 AC ADN26956;
 XX
 DT 02-DEC-2004 (first entry)
 XX
 DE Bacterial polypeptide #9609.
 XX
 KW Recombinant DNA construct; transformed plant; improved plant property;

KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polypeptide.
 XX Bacteria.
 XX OS
 XX US2003233675-A1.
 XX 18-DEC-2003.
 XX PD
 XX PF 20-FEB-2003; 2003US-00369493.
 XX PR 21-FEB-2002; 2002US-0360039P.
 XX PA (CAOY/) CAO Y.
 XX PA (HINK/) HINKLE G J.
 XX PA (SLAT/) SLATER S C.
 XX PA (CHEN/) CHEN X.
 XX PA (GOLD/) GOLDMAN B S.
 XX PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 XX WPI; 2004-061375/06.
 XX DR
 XX PT New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.
 XX PS Claim 1; SEQ ID NO 9609; 122pp; English.
 XX CC The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polypeptide used in the
 CC scope of the invention. Note: The sequence data for this patent did not
 CC form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.
 XX SQ Sequence 786 AA;
 SQ Query Match 46.1%; Score 47; DB 8; Length 786;
 Best Local Similarity 47.4%; Pred. NO. 40;
 Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 QY 5 PFSVKDILSLPQXRXAGA 23
 DB 345 PLTVEDVLSYKERPIGA 363
 RESULT 24
 AAR83015
 ID AAR83015 standard; protein; 371 AA.
 XX

AC AAR83015;
 XX 27-NOV-1996 (first entry)
 DT Human thyroid transcription factor-1.
 XX DE
 XX KW Thyroid transcription factor; TTF-1; human adenocarcinoma cell line;
 KW H441; rat; mouse; pulmonary adenocarcinoma; H820; small cell carcinoma;
 KW H345; tracheal-bronchial epithelial cell lines; respiratory epithelium;
 KW fetal lung; gestation; pro-SP-C; respiratory epithelial cell;
 KW nonciliated bronchiolar cell; immature lung; alveolar; Type II;
 KW epithelial cell; nonciliated; Type I; adult.
 XX OS Homo sapiens.
 XX W09531729-A1.
 XX 23-NOV-1995.
 PD 17-MAY-1995; 95WO-US006244.
 PF 18-MAY-1994; 94US-00245356.
 PR (CHIL-) CHILDRENS HOSPITAL MEDICAL CENT.
 XX PA Bohinski RJ, Whitsett JA;
 XX PI WPI; 1996-011078/01.
 XX DR N-PSDB; AAT05810.
 XX PT Oligo- and polynucleotide(s) that bind to lung cell nuclear proteins -
 PT useful for cancer diagnosis and therapy.
 XX PS Claim 50; Fig 39; 157pp; English.
 XX CC This sequence represents the thyroid transcription factor, TTF-1. The TTF
 CC -1 DNA sequence was isolated from the human adenocarcinoma cell line
 CC H441. The TTF-1 locus is contained within a 4.6 kb BamHI fragment and
 CC consists of two exons and one intron. This predicted amino acid sequence
 CC of human TTF-1 shows close identity with the amino acid sequence
 CC predicted by the rat TTF-1 cDNA sequence and 92.4% identity with the rat
 CC TTF-1 cDNA. The intron is approx. 1 kb in length and is flanked by
 CC consensus splice donor-acceptor sites that fit splice acceptor-donor
 CC rules. The mRNA produced is 2.3 kb as detected by Northern blot analysis
 CC of mRNA derived from mouse and human adenocarcinoma cells. TTF-1 mRNA was
 CC detected in human pulmonary adenocarcinoma cells H441 and H820 and small
 CC cell carcinoma H345, but not detected in 9/HTEO- or BEAS-2B tracheal-
 CC bronchial epithelial cell lines, A549, HeLa or 3T3 cells, demonstrating
 CC the cell selectivity of TTF-1 expression. TTF-1 has been detected in
 CC nuclei of the respiratory epithelium in human fetal lungs as early as 11-
 CC 12 weeks of gestation. Immunostaining has observed a distribution pattern
 CC in the developing airways similar to the for pro-SP-C. TTF-1 was detected
 CC in subsets of respiratory epithelial cells in the developing lung,
 CC including nonciliated bronchiolar, and rarely in nonciliated bronchila
 CC respiratory epithelial cells in the immature lung. At the time of birth
 CC TTF-1 was detected in alveolar Type II epithelial cells and in subsets of
 CC nonciliated bronchiolar epithelial cells. TTF-1 was not detected in
 CC alveolar Type I cells or ciliated epithelial cells. In the adult lung,
 CC TTF-1 is detected in subsets of nonciliated bronchiolar epithelial cells
 CC and was most prominent in type II epithelial cells but was excluded from
 CC Type I cells
 XX SQ Sequence 371 AA;
 SQ Query Match 45.6%; Score 46.5; DB 2; Length 371;
 Best Local Similarity 60.0%; Pred. No. 21;
 Matches 12; Conservative 1; Mismatches 6; Indels 1; Gaps 1;
 QY 4 TPFSVKDILSLPQXRXAG 22
 DB 9 TPFSVSDILSPLEESYKKVG 28


```

RESULT 25
AAU05583
ID AAU05583 standard; protein; 371 AA.
XX AC
XX AAU05583;
XX 24-OCT-2001 (first entry)
XX DT
XX Human Thyroid transcription factor.
XX DE
XX Human; thyroid transcription factor; TTF-1; lung cancer; thyroid cancer.
XX KW
XX Homo sapiens.
XX OS
XX US2001016352-A1.
XX PN
XX 23-AUG-2001.
XX PD
XX 26-MAY-1999; 99US-00320337.
XX PF
XX 18-MAY-1994; 94US-00245336.
XX PR
XX 17-MAY-1995; 95US-00442809.
XX PR
XX (BOHI/) BOHINSKI R J.
XX PA (WHIT/) WHITSETT J A.
XX PA
XX Bohinski RJ, Whitsett JA;
XX PI
XX WPI; 2001-513959/56.
XX DR N-PSDB; AAS11829.
XX DR
XX Oligonucleotide sequences which bind nuclear proteins and surfactants
XX PT found in lung cells, useful for detecting cancers that originate in the
XX PT lung.
XX PT
XX Example 7; Fig 39; 76pp; English.
XX PS
XX The invention relates to an oligonucleotide which includes at least 1
XX CC nucleic acid sequence which binds to at least 1 nuclear protein found in
XX CC lung cells (e.g. the thyroid transcription factor 1, TTF-1, protein). The
XX CC oligonucleotide can be expressed in lung cells via a vector and can be
XX CC used to target therapeutic agents to kill lung or thyroid cancer cells.
XX CC The oligonucleotide can be used to detect or diagnose lung or thyroid
XX CC cancer. The oligonucleotides may be designed from the sequences of, for
XX CC example, the promoters of lung-specific genes such as those encoding
XX CC surfactant proteins. The present sequence is the human thyroid
XX CC transcription factor, TTF-1
XX CC
XX CC Sequence 371 AA;
XX SQ
Query Match 45.6%; Score 46.5; DB 4; Length 371;
Best Local Similarity 60.0%; Pred. No. 21;
Matches 12; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 4 TPFSVKDILS-LPEQXRANG 22
DB 9 TPFSVSDILSPLSESYKKVG 28
||||| ||||| :
TPFSVSDILSPLSESYKKVG 28

RESULT 26
ADB70334
ID ADB70334 standard; protein; 401 AA.
XX AC
XX ADB70334;
XX AC
XX 04-DEC-2003 (first entry)
XX DT
XX Thyroid transcription factor 1 SEQ ID NO:26.
XX DE
XX cancer; malignant pleural mesothelioma; MPN; lung adenocarcinoma;
XX KW squamous carcinoma; medulloblastoma; prostate cancer; breast cancer;
XX KW diffuse large B-cell lymphoma; follicular lymphoma; ovarian cancer;
XX KW human.

RESULT 27
ADJ37113
ID ADJ37113 standard; protein; 401 AA.
XX AC
XX ADJ37113;
XX AC
XX 22-APR-2004 (first entry)
XX DT
XX Human malignant pleural mesothelioma (MPN) protein #9.
XX DE
XX Human; malignant pleural mesothelioma; MPN; tumour; lung adenocarcinoma;
XX KW squamous carcinoma; medulloblastoma; prostate cancer; breast cancer;
XX KW diffuse large B-cell lymphoma; follicular lymphoma; ovarian cancer;
XX KW cytostatic.
XX XX
XX Homo sapiens.
XX OS
XX US2003219760-A1.
XX PN
XX

```

```

XX Homo sapiens.
XX OS
XX WO2003021229-A2.
XX PN
XX 13-MAR-2003.
XX PD
XX 05-SEP-2002; 2002WO-US028203.
XX PF
XX 05-SEP-2001; 2001US-0317389P.
XX PR
XX 30-AUG-2002; 2002US-00236031.
XX PR
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.
XX PA
XX Gordon GJ, Jensen RV, Gullans SR, Bueno R;
XX PI WPI; 2003-290233/28.
XX DR N-PSDB; ADB70333.
XX DR
XX Diagnosing cancer cells in tissue sample, or determining prognosis or,
XX PT outcome of cancer patient, by calculating ratio of expression levels of
XX PT genes that are differentially expressed in cancer and non cancer tissues.
XX PT
XX Claim 77; Page 142-143; 396pp; English.
XX PS
XX The present invention describes a method (M1) for diagnosing the presence
XX CC of cancer cells or non-cancer cells in a tissue sample, or determining
XX CC the prognosis or outcome of a cancer patient. M1 involves providing a set
XX CC of genes that are differentially expressed in cancerous or non-cancerous
XX CC conditions, determining the expression levels of the set of genes and
XX CC calculating a ratio of the expression levels of the differentially
XX CC expressed genes. M1 is useful for diagnosing the presence of cancer cells
XX CC or non-cancer cells in a tissue sample, where the cancer is malignant
XX CC pleural mesothelioma (MPM), lung adenocarcinoma, squamous carcinoma,
XX CC medulloblastoma, prostate cancer, breast cancer, diffuse large B-cell
XX CC lymphoma, follicular lymphoma and ovarian cancer, and for determining
XX CC prognosis or outcome of a cancer patient. The ratio of expression levels
XX CC of differentially expressed genes is used as an indicator of cancer type,
XX CC cancer class, and/or cancer prognosis, all of which are useful for
XX CC determining a course of treatment of a patient. The present sequence
XX CC represents a human protein which is used in an example from the present
XX CC invention.
XX CC
XX CC Sequence 401 AA;
XX SQ
Query Match 45.6%; Score 46.5; DB 7; Length 401;
Best Local Similarity 60.0%; Pred. No. 23;
Matches 12; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 4 TPFSVKDILS-LPEQXRANG 22
DB 39 TPFSVSDILSPLSESYKKVG 58
||||| ||||| :
TPFSVSDILSPLSESYKKVG 58

RESULT 27
ADJ37113
ID ADJ37113 standard; protein; 401 AA.
XX AC
XX ADJ37113;
XX AC
XX 22-APR-2004 (first entry)
XX DT
XX Human malignant pleural mesothelioma (MPN) protein #9.
XX DE
XX Human; malignant pleural mesothelioma; MPN; tumour; lung adenocarcinoma;
XX KW squamous carcinoma; medulloblastoma; prostate cancer; breast cancer;
XX KW diffuse large B-cell lymphoma; follicular lymphoma; ovarian cancer;
XX KW cytostatic.
XX XX
XX Homo sapiens.
XX OS
XX US2003219760-A1.
XX PN
XX

```


XX Ericson J;
XX WPI; 2002-657380/70.
XX
XX Guiding fate of differentiation of cell into cell type by contacting cell
XX with Groucho-interacting protein to form a complex which represses DNA
XX transcription and suppresses alternative pathways of differentiation.
XX
XX Example 2; Fig 1; 116pp; English.
XX
XX The present invention describes a method (M1) for guiding the fate of
XX differentiation of a cell into a specific cell type, by providing a
XX sample comprising the cell, contacting the sample with a Groucho-
XX interacting protein (GIP) to result in the formation of a complex between
XX GIP and a Groucho corepressor protein (I), where the complex represses
XX DNA transcription and suppresses alternative pathways of differentiation.
XX A purified complex (PC1) comprising a first polypeptide comprising a GIP
XX and the second polypeptide comprising a Groucho corepressor protein, can
XX be used for identifying an agent which modulates the stability or
XX activity of the complex, or that disrupts a polypeptide complex. PC1 is
XX useful for screening the candidate substance interacting with the complex
XX or with (I); for inhibiting the guided differentiation of a cell
XX resulting in the impairment of ventral patterning; for identifying a
XX polypeptide complex in a subject; and determining altered expression of a
XX polypeptide in a subject. M1 is useful for guiding the fate of
XX differentiation of a cell into a specific cell type. The present sequence
XX represents a GIP related peptide, which is used in the exemplification of
XX the present invention
XX
XX Sequence 23 AA;
XX
XX Query Match 45.1%; Score 46; DB 5; Length 23;
XX Best Local Similarity 50.0%; Pred. NO. 1;
XX Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
Qy 4 TPFSVKDILSLPEQXRAXGA 23
Db 4 TPFGINDILSRPSPVAGSA 23
RESULT 29
ADQ09852
ID ADQ09852 standard; protein; 365 AA.
XX
XX ADQ09852;
XX
XX 23-SEP-2004 (first entry)
XX
XX Mouse NK-2 class homeobox protein, NKx6.1.
XX
XX Mouse; islet cell differentiation transcription factor;
XX insulin-dependent diabetes; insulin; somatic cell;
XX insulin-producing cell; islet cell; NeuroD; ngn3; neurogenin 3; Pax6;
XX paired-box transcription factor 6; Pax4;
XX paired-box transcription factor 4; Nkx2.2; NK-2 class homeobox protein;
XX Nkx6.1; Isl-1; Islet factor 1; Pdx-1;
XX pancreatic and duodenal homeobox protein; BCT; betacellulin.
XX
XX Mus musculus.
XX
XX US2004132679-A1.
XX
XX 08-JUL-2004.
XX
XX 03-SEP-2003; 2003US-00654102.
XX
XX 03-SEP-2002; 2002US-0407743P.
XX
XX (BAYU) BAYLOR COLLEGE MEDICINE.
XX
XX Chan L, Kojima H;
XX
XX

PD 27-NOV-2003.
XX
XX 05-SEP-2002; 2002US-00236031.
XX
XX 05-SEP-2001; 2001US-0317389P.
XX
XX 30-AUG-2002; 2002US-0407431P.
XX
XX (BGHM) BRIGHAM & WOMENS HOSPITAL INC.
XX
XX Gordon GJ, Jensen RV, Gullans SR, Bueno R;
XX
XX WPI; 2004-141744/14.
XX
XX N-PSDB; ADJ37112.
XX
XX Diagnosing the presence of cancer or non-cancer cells in tissue sample,
XX useful for diagnosing malignant pleural mesothelioma comprises
XX determining ratio of expression level of a set of genes expressed in
XX cancer tissues.
XX
XX Claim 77; SEQ ID NO 26; 53pp; English.
XX
XX The invention relates to a method of diagnosing the presence of cancer
XX cells or non-cancer cells in a tissue sample, determining prognosis or
XX outcome of a cancer patient, selecting a course of treatment for a
XX subject having or suspected of having malignant pleural mesothelioma
XX (MPM) and evaluating treatment of MPM comprising determining the ratio of
XX the expression level of a set of genes differentially expressed in a
XX cancer tissue. The cancer is chosen from MPM, lung adenocarcinoma,
XX squamous carcinoma, medulloblastoma, prostate cancer, breast cancer,
XX diffuse large B-cell lymphoma, follicular lymphoma and ovarian cancer.
XX The method is useful for diagnosing MPM in a subject suspected of having
XX MPM which involves obtaining a tissue sample suspected of being cancerous
XX from a subject and determining the expression of nucleic acid markers or
XX its expression products in the tissue sample. This sequence represents a
XX human MPM protein of the invention. Note: The sequence data for this
XX patent did not form part of the printed specification but was obtained in
XX electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX Sequence 401 AA;
XX
XX Query Match 45.6%; Score 46.5; DB 8; Length 401;
XX Best Local Similarity 60.0%; Pred. NO. 23;
XX Matches 12; Conservative 1; Mismatches 6; Indels 1; Gaps 1;
Qy 4 TPFSVKDILSLPEQXRAXG 22
Db 39 TPFSVSDILSLPEESYKVG 58
RESULT 28
ABP53050
ID ABP53050 standard; peptide; 23 AA.
XX
XX ABP53050;
XX
XX 06-NOV-2002 (first entry)
XX
XX GIP related peptide rNkx6.1 SEQ ID NO:3.
XX
XX Human; Groucho-interacting protein; GIP; Nkx; Nkx6.3; differentiation;
XX Groucho corepressor protein.
XX
XX Rattus sp.
XX
XX WO200242413-A2.
XX
XX 30-MAY-2002.
XX
XX 01-NOV-2001; 2001WO-IB002835.
XX
XX 01-NOV-2000; 2000US-0245002P.
XX
XX (NEUR-) NEURONOVA AB.
XX
XX

DR WPI: 2004-517032/49.
 DR N-PSDB; ADQ09873, ADQ09875, ADQ09947.
 XX
 PT Use of an islet cell differentiation transcription factor polypeptide or
 PT its homologue or analog for treating a mammal for insulin-dependent
 PT diabetes, increasing an insulin level in a somatic cell, or generating an
 PT insulin-producing cell.
 XX
 PS Claim 109; SEQ ID NO 90; 190pp; English.
 XX
 CC The invention relates to the use of an islet cell differentiation
 CC transcription factor polypeptide or its homologue or analogue for
 CC treating a mammal for insulin-dependent diabetes, increasing an insulin
 CC level in a somatic cell, or generating an insulin-producing cell. Also
 CC included are a method of treating a mammal for insulin-dependent
 CC diabetes, a method of increasing an insulin level in a somatic cell, a
 CC method of generating an insulin-producing cell, a therapeutic composition
 CC comprising an isolated islet cell differentiation transcription factor
 CC polypeptide (and/or an isolated nucleic acid expressing the polypeptide),
 CC an insulin-producing cell comprising a vector (where the vector comprises
 CC a nucleic acid sequence encoding an islet cell differentiation
 CC transcription factor), an insulin-producing cell (generated by a method
 CC comprising obtaining a somatic cell and transfecting the cell with a
 CC vector comprising a nucleic acid sequence encoding an islet cell
 CC differentiation transcription factor, where in the transfecting step the
 CC cell produces insulin), a method of generating at least one pancreatic
 CC islet, and a composition (comprising: NeuroD or ng3 polypeptide or a
 CC polynucleotide expressing a NeuroD or ng3 polypeptide and betacellulin
 CC polypeptide or a polynucleotide expressing a betacellulin polypeptide).
 CC The islet cell differentiation transcription factor polypeptide is
 CC NeuroD, ng3 (neurogenin 3), Pax6 (paired-box transcription factor 6),
 CC Pax4 (paired-box transcription factor 4), Nkx2.2 (NK-2 class homeobox
 CC protein), Nkx6.1, Isl-1 (islet factor 1), Pdx-1 (pancreatic and duodenal
 CC homeobox protein), BCT (betacellulin) or their combinations. The islet
 CC cell differentiation transcription factor polypeptide or its homologue or
 CC analogue is useful for treating a mammal for insulin-dependent diabetes,
 CC increasing an insulin level in a somatic cell, or generating an insulin-
 CC producing cell. The present sequence is an islet cell differentiation
 CC transcription factor polypeptide as detailed above.
 XX
 SQ Sequence 365 AA;
 Query Match 45.1%; Score 46; DB 8; Length 365;
 Best Local Similarity 50.0%; Pred. No. 25;
 Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
 QY 4 TPFSVKDILSLPEQXAXGA 23
 || : ||| : ||
 DB 94 TPGINDILSRPMPVAGSA 113
 || : ||| : ||
 RESULT 30
 ABG19376
 ID ABG19376 standard; protein; 171 AA.
 XX
 AC ABG19376;
 XX
 DT 18-FEB-2002 (first entry)
 DE
 DE Novel human diagnostic protein #19367.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 XX

PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR N-PSDB; AAS83563.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 49735; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 171 AA;
 Query Match 44.1%; Score 45; DB 4; Length 171;
 Best Local Similarity 61.5%; Pred. No. 16;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 5 PFSVKDILSLPEQ 17
 || : ||| : :
 DB 65 PFKIKDILSLKKE 77
 || : ||| : :
 RESULT 31
 ABM67666
 ID ABM67666 standard; protein; 321 AA.
 XX
 AC ABM67666;
 XX
 DT 20-NOV-2003 (first entry)
 DE
 DE Photorhabdus luminescens protein sequence #763.
 XX
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 KW whooping cough.
 XX
 OS Photorhabdus luminescens.
 XX
 PN WO200294867-A2.
 XX
 PD 28-NOV-2002.
 XX
 PF 07-FEB-2002; 2002WO-IB003040.
 XX
 PR 07-FEB-2001; 2001FR-00001659.
 XX

XX	(INSP) INST PASTEUR.
PA	(CNRS) CNRS CENT NAT RECH SCI.
XX	Duchaud E, Taourit S, Glaser P, Frangeul L, Kunet F, Danchin A;
PI	Buchrieser C;
XX	WPI; 2003-148459/14.
DR	
XX	Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT	useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX	Claim 2; SEQ ID NO 763; 1205pp; French.
PS	
XX	The invention relates to the isolation of genes and their encoded
CC	proteins from Photorhabdus luminescens. The isolated sequences are
CC	sources of probes and primers for detecting the genome of P. luminescens
CC	and related species; to study polymorphisms; for gene analysis and for
CC	detection/amplification of the genes. Antibodies (Ab) raised against the
CC	polypeptides encoded by the genes are used for detection/identification
CC	of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC	carry a gene-containing vector are used to select compounds that
CC	modulate, regulate, induce or inhibit expression of the genes in plants,
CC	animals or microorganisms other than P. luminescens and are able to alter
CC	response or sensitivity to toxins and antibiotics produced by P.
CC	luminescens. Cells transformed to express the genes are useful for
CC	recombinant production of the proteins, particularly toxins and
CC	antibacterials useful as insecticides, bactericides and fungicides. The
CC	genes, proteins, vectors containing the genes and Ab are also useful
CC	therapeutically (to treat microbial infection by bacteria or fungi that
CC	are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC	biopesticides. Other uses of the genes and the proteins are as virulence
CC	factors and for identifying targets of human diseases for which P.
CC	luminescens is a model (particularly plague and whooping cough). This
CC	sequence represents one of the isolated P. luminescens proteins
XX	
SQ	Sequence 321 AA;
	Query Match 44.1%; Score 45; DB 6; Length 321;
	Best Local Similarity 50.0%; Pred. No. 32;
	Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
OY	4 TPFSSVKDILSLPEQ 17
DB	: : :
	303 TFFNTETIIISLPRE 316
RESULT 32	
ADF06776	
ID	ADF06776 standard; protein; 338 AA.
XX	AC
AD	ADF06776;
XX	DT
DE	12-FEB-2004 (first entry)
XX	DE
XX	Bacterial polypeptide #2889.
KW	Proteus mirabilis infection; bacterial infection, antibacterial;
KW	immunostimulant.
OS	Proteus mirabilis.
XX	XX
PN	US6605709-B1.
XX	PD
XX	12-AUG-2003.
XX	05-APR-2000; 2000US-00543681.
XX	PR
XX	09-APR-1999; 99US-0128706P.
XX	(GENO-) GENOME THERAPEUTICS CORP.
PA	
PI	Breton'GL;

XX
DR DR WIPI; 2001-895291/82.
XX N-PSDB; ADF02604.

XX
PT PT New Proteus mirabilis polypeptides and polynucleotides, useful as
PT reagents for diagnosis of bacterial disease, as components of
PT antibacterial vaccines, as targets for antibacterial drugs, or as
PT biocontrol agents for plants.

XX
XX Disclosure; SEQ ID NO 7061; 870pp; English.

XX
CC The invention relates to new Proteus mirabilis polypeptides and
CC polynucleotides. The invention also relates to antibodies against the
CC polypeptides. Methods for producing the polypeptides, a method of
CC generating vaccines for immunising an individual against P. mirabilis,
CC method for evaluating a compound for the ability to bind a P. mirabilis
CC polypeptide and a method for screening test compounds for anti-bacterial
CC activity. The polypeptides and polynucleotides are useful as molecular
CC targets for diagnosing, preventing and treating pathological conditions
CC resulting from bacterial infection, as reagents for diagnosis of
CC bacterial diseases, as components of antibacterial vaccines, as targets
CC for antibacterial drugs or as bio-control agents for plants. This
CC sequence represents a Proteus mirabilis polypeptide of the invention.

XX
SQ Sequence 338 AA;

Query Match 44.1%; Score 45; DB 7; Length 338;
Best Local Similarity 52.6%; Pred. No. 34;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0

OY 4 TPFSVKDILSLPEQXRAXG 22
||||| :||| |
DB 81 TPFSVEMILSIERTARQFG 99

RESULT 33
ABG04855
ID ABG04855 standard; protein; 197 AA.
XX AC ABG04855;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #4846.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX WO200175067-A2.
PN PN
PD PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WIPI; 2001-639362/73.
DR N-PSDB; AAS69042.

XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX
PS Claim 20; SEQ ID NO 35214; 103pp; English.
XX

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 197 AA;

Query Match 43.1%; Score 44; DB 4; Length 197;
 Best Local Similarity 72.7%; Pred. No. 28;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 TKTPTSVKDL 12
 Db 56 TSTPTLKDIL 66

RESULT 34
 ADN99762
 ID ADN99762 standard; protein; 322 AA.

XX AC ADN99762;

XX DT 29-JUL-2004 (first entry)

XX DE Novel human protein sequence #578.

XX KW anti-inflammatory; dermatological; neuroprotective; immunomodulator;
 KW antibacterial; virucide; antiparasitic; cytostatic; gene therapy;
 KW vaccine; inflammatory; CNS; immune disorder; cancer; psoriasis; diabetes;
 KW early aging; hormonal imbalance; ischemic heart disease;
 KW ulcerative colitis.

XX OS Homo sapiens.

XX PN WO2004038003-A2.

XX PD 06-MAY-2004.

XX PF 24-OCT-2003; 2003WO-US033947.

XX PR 25-OCT-2002; 2002US-0421061P.

PR 25-OCT-2002; 2002US-0421080P.

PR 25-OCT-2002; 2002US-0421552P.

PR 25-OCT-2002; 2002US-0421614P.

PR 30-OCT-2002; 2002US-0422177P.

PR 30-OCT-2002; 2002US-0422178P.

PR 15-NOV-2002; 2002US-0426355P.

PR 15-NOV-2002; 2002US-0426384P.

PR 15-NOV-2002; 2002US-0426430P.

PR 15-NOV-2002; 2002US-0426916P.

PR 27-NOV-2002; 2002US-0429224P.

PR 27-NOV-2002; 2002US-0429275P.

PR 27-NOV-2002; 2002US-0429302P.

PR 27-NOV-2002; 2002US-0429326P.

PR 27-NOV-2002; 2002US-0429651P.

PR 04-DEC-2002; 2002US-0430645P.
 PR 04-DEC-2002; 2002US-0430651P.
 PR 04-DEC-2002; 2002US-0430657P.
 PR 04-DEC-2002; 2002US-0430663P.
 PR 04-DEC-2002; 2002US-0430668P.
 PR 04-DEC-2002; 2002US-0430684P.
 PR 05-DEC-2002; 2002US-0430937P.
 PR 05-DEC-2002; 2002US-0430965P.
 PR 05-DEC-2002; 2002US-0431458P.
 PR 12-DEC-2002; 2002US-0433251P.
 PR 12-DEC-2002; 2002US-0433500P.
 PR 13-DEC-2002; 2002US-0433316P.
 PR 13-DEC-2002; 2002US-0433318P.
 PR 13-DEC-2002; 2002US-0436238P.
 PR 03-JAN-2003; 2003US-0437914P.
 PR 17-JAN-2003; 2003US-0440820P.
 PR 17-JAN-2003; 2003US-0440821P.
 PR 18-APR-2003; 2003US-0463700P.
 PR 18-APR-2003; 2003US-0463708P.
 PR 18-APR-2003; 2003US-0463718P.
 PR 18-APR-2003; 2003US-0463732P.
 PR 02-MAY-2003; 2003US-0467199P.
 PR 02-MAY-2003; 2003US-0467201P.
 PR 02-MAY-2003; 2003US-0467203P.
 PR 02-MAY-2003; 2003US-0467230P.
 PR 19-MAY-2003; 2003US-0471308P.
 PR 19-MAY-2003; 2003US-0471336P.
 PR 22-MAY-2003; 2003US-0472420P.
 PR 22-MAY-2003; 2003US-0472430P.
 PR 09-JUN-2003; 2003US-0476609P.
 PR 09-JUN-2003; 2003US-0476621P.
 PR 09-JUN-2003; 2003US-0476632P.
 PR 09-JUN-2003; 2003US-0476641P.
 PR 08-JUL-2003; 2003US-0485217P.
 PR 08-JUL-2003; 2003US-0485218P.
 PR 08-JUL-2003; 2003US-0485223P.
 PR 08-JUL-2003; 2003US-0485224P.
 PR 08-JUL-2003; 2003US-0485325P.
 PR 08-JUL-2003; 2003US-0485359P.
 PR 14-JUL-2003; 2003US-0486446P.
 PR 14-JUL-2003; 2003US-0486480P.
 PR 15-JUL-2003; 2003US-0486891P.
 PR 15-JUL-2003; 2003US-0486960P.
 PR 08-AUG-2003; 2003US-0493341P.
 PR 08-AUG-2003; 2003US-0493370P.
 PR 08-AUG-2003; 2003US-0493573P.
 PR 08-AUG-2003; 2003US-0493577P.

(FIVE-) FIVE PRIME THERAPEUTICS INC.

XX Williams LT, Chu K, Lee B, Heatir K, Beaurang PA, Behrens D;
 PI Halenbeck RF, Kothakota S, Lin H, Linnemann T, Pierce K, Wang Y;
 PI Wong JGP, Wu G, Zhang H, Zeng C;

DR WPI; 2004-365511/34.

DR N-PSDB; ADN98978.

XX New nucleic acid molecules, useful in preparing a composition for
 PT treating or preventing e.g. inflammatory, CNS, bacterial or viral
 PT disorders, cancer, psoriasis, diabetes, ischemic heart disease or
 PT ulcerative colitis.

PS Claim 14; SEQ ID NO 1362; 532pp; English.

CC The invention relates to a nucleic acid molecule comprising a
 CC polynucleotide sequence or its complement that encodes a polypeptide. The
 CC nucleic acid is useful in preparing a composition for treating or
 CC preventing inflammatory, CNS, immune, bacterial or viral disorder,
 CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
 CC heart disease or ulcerative colitis. This sequence corresponds to a
 CC protein of the invention.

XX Sequence 322 AA;

XX Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
KW Drosophila melanogaster.
XX WO200171042-A2.
OS 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX N-PSDB; ABL07678.
DR New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.
XX Disclosure; SEQ ID NO 17517; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB57072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 382 AA;
SQ Query Match 43.1%; Score 44; DB 4; Length 382;
Best Local Similarity 61.1%; Pred. No. 59;
Matches 11; Conservative 1; Mismatches 4; Indels 2; Gaps 1;
QY 4 TTFPSVKDIL--SLPEQXR 19
Db 22 TTFPSINDILTRSNPETR 39
RESULT 37
ABB04859
ID ABB04859 standard; protein; 471 AA.
XX AC ABB04859;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #4850.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
PR

XX Query Match 43.1%; Score 44; DB 8; Length 322;
KW Best Local Similarity 72.7%; Pred. No. 49;
XX Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 TKTTFPSVKDIL 12
Db 52 TSTPFTLKDIL 62
RESULT 35
ABB09760
ID ABB09760 standard; protein; 349 AA.
XX AC ABB09760;
XX DT 12-JUN-2002 (first entry)
XX DE Amino acid sequence of human hearing defect related protein 38.39.
XX KW Human; hearing defect related protein 38.39; presbycusis;
KW non integrated hearing defect; congenital deafness; auricle abnormality;
KW gene therapy.
XX OS Homo sapiens.
XX WO200220606-A1.
XX 14-MAR-2002.
XX 03-SEP-2001; 2001WO-CN001324.
XX 07-SEP-2000; 2000CN-00125066.
XX (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.
XX MAO Y, Xie Y;
XX WPI; 2002-292256/33.
XX N-PSDB; ABL53170.
XX Human hearing defect related protein 38.39 and encoding polynucleotide, useful in the treatment of e.g., presbycusis, non integrated hearing defect, congenital deafness and auricle abnormality.
XX Claim 1; Page 27-28; 37pp; Chinese.
XX The present sequence represents human hearing defect related protein 38.39. The human hearing-defect related protein 38.39 and its encoding polynucleotide are useful in the treatment of e.g., presbycusis, non integrated hearing defect, congenital deafness and auricle abnormality. The polynucleotide is also used for gene therapy
XX Sequence 349 AA;
SQ Query Match 43.1%; Score 44; DB 5; Length 349;
Best Local Similarity 72.7%; Pred. No. 53;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 TKTTFPSVKDIL 12
Db 49 TSTPFTLKDIL 59
RESULT 36
ABB63575
ID ABB63575 standard; protein; 382 AA.
XX AC ABB63575;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 17517.


```

XX PA (HYSE-) HYSEQ INC.
XX FI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS69046.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 35218; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG0010-ABG3037 represent novel human diagnostic
XX CC amino acid sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 471 AA;

Query Match 43.1%; Score 44; DB 4; Length 471;
Best Local Similarity 72.7%; Pred. No. 75;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 TKTPTSVKDIIL 12
Db 5 TSTPTLTLDIL 15

RESULT 38
AAW03528
ID AAW03528 standard; protein; 553 AA.
XX AC AAW03528;
XX DT 25-FEB-1997 (first entry)
XX DE Salmonella enterica IagA protein.
XX KW IagA; IagB; Salmonella enterica; primer; probe; HeLa.
XX OS Synthetic.
XX PN EP721989-A1.
XX PD 17-JUL-1996.
XX PF 15-JAN-1996; 96EP-00400098.
XX PR 16-JAN-1995; 95FR-00000410.
XX PA (INSP ) INST PASTEUR.
XX PA (INRM ) INSERM NAT SANTE & RECH MEDICALE.
XX

PI Popoff MY, Le Guern Fellous M;
DR WPI; 1996-322837/33.
DR N-PSDB; AAT37466.
XX PT New nucleic acid of S. enterica spp. enterica involved in cell invasion -
XX PT and derived oligo-nucleotide(s) useful as primers and probes for
XX PT detecting Salmonella in food etc.
XX PS Claim 13; Fig 1; 33pp; French.
XX CC This is the amino acid sequence of the IagA protein from Salmonella
XX CC enterica spp. enterica serovar typhi. The protein has a calculated mol.
XX CC wt. of 63026 Da. The N-terminal portion of the protein has homology with
XX CC a similar region of the transcriptional regulatory proteins PhoB (24%
XX CC identity and 52% similarity) and PhoP (25% identity and 62% similarity)
XX CC from E. coli. The IagA and IagB sequences are used to generate primers and
XX CC probes (AAT37451-65) which are useful for detecting S. enterica and/or S.
XX CC bongori in samples e.g. in food. The IagA and IagB proteins are involved
XX CC in invasion of cultured HeLa cells by S. enterica
XX SQ Sequence 553 AA;

Query Match 43.1%; Score 44; DB 2; Length 553;
Best Local Similarity 61.1%; Pred. No. 90;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 TKTPTSVKDIILSLPEQXR 19
Db 164 TKNCRSVKDILSLMDQLR 181

RESULT 39
ADN21330
ID ADN21330 standard; protein; 732 AA.
XX AC ADN21330;
XX DT 02-DEC-2004 (first entry)
XX DE Bacterial polypeptide #3983.
XX KW Recombinant DNA construct; transformed plant; improved plant property;
XX KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
XX KW pathogen tolerance; pest tolerance; plant disease resistance;
XX KW cell cycle pathway modification; plant growth regulator;
XX KW homologous recombination; seed oil yield; protein yield; carbohydrate;
XX KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
XX KW bacterial polypeptide.
XX OS Bacteria.
XX PN US2003233675-A1.
XX PD 18-DEC-2003.
XX PF 20-FEB-2003; 2003US-00369493.
XX PR 21-FEB-2002; 2002US-0360039P.
XX PA (CAOY/) CAO Y.
XX PA (HINK/) HINKLE G J.
XX PA (SLAT/) SLATER S C.
XX PA (CHEN/) CHEN X.
XX PA (GOLD/) GOLDMAN B S.
XX PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX DR WPI; 2004-061375/06.
XX PT New recombinant DNA construct comprising a promoter positioned to provide
XX PT for expression of a polynucleotide encoding a polypeptide from a
XX PT microbial source, useful for producing plants with improved properties.

```


XX Use of an islet cell differentiation transcription factor polypeptide or
PT its homologue or analog for treating a mammal for insulin-dependent
PT diabetes, increasing an insulin level in a somatic cell, or generating an
PT insulin-producing cell.
XX Claim 109; SEQ ID NO 88; 190pp; English.
XX The invention relates to the use of an islet cell differentiation
CC transcription factor polypeptide or its homologue or analogue for
CC treating a mammal for insulin-dependent diabetes, increasing an insulin
CC level in a somatic cell, or generating an insulin-producing cell. Also
CC included are a method of treating a mammal for insulin-dependent
CC diabetes, a method of increasing an insulin level in a somatic cell, a
CC method of generating an insulin-producing cell, a therapeutic composition
CC comprising an isolated islet cell differentiation transcription factor
CC polypeptide (and/or an isolated nucleic acid expressing the polypeptide),
CC an insulin-producing cell comprising an islet cell differentiation
CC a nucleic acid sequence encoding an islet cell differentiation
CC transcription factor), an insulin-producing cell (generated by a method
CC vector comprising a nucleic acid sequence encoding an islet cell
CC differentiation transcription factor, where in the transfecting step the
CC cell produces insulin), a method of generating at least one pancreatic
CC islet, and a composition (comprising: NeuroD or ngn3 polypeptide or a
CC polynucleotide expressing a NeuroD or ngn3 polypeptide and betacellulin
CC polypeptide or a polynucleotide expressing a betacellulin polypeptide).
CC The islet cell differentiation transcription factor polypeptide is
CC NeuroD, ngn3 (neurogenin 3), Pax6 (paired-box transcription factor 6),
CC Pax4 (paired-box transcription factor 4), Nkx2.2 (NK-2 class homeobox
CC protein), Isl-1 (islet factor 1), Pdx-1 (pancreatic and duodenal
CC homeobox protein), BCT (betacellulin) or their combinations. The islet
CC cell differentiation transcription factor polypeptide or its homologue or
CC analogue is useful for treating a mammal for insulin-dependent diabetes,
CC increasing an insulin level in a somatic cell, or generating an insulin-
CC producing cell. The present sequence is an islet cell differentiation
CC transcription factor polypeptide as detailed above.
XX SQ Sequence 87 AA;

Query Match 42.2%; Score 43; DB 8; Length 87;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 6 FSVKXDLISLPEQXAXGA 23
Db 11 FKVKDILDLPTDNDGGS 28

Search completed: April 13, 2005, 17:11:25
Job time : 26.6993 secs

XX The invention relates to a recombinant DNA construct comprising a
PS promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plants with
CC The recombinant DNA construct is useful for producing plants with
CC tolerance to herbicides, e.g. improved cold, heat or drought tolerance,
CC improved resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX SQ Sequence 732 AA;

Query Match 43.1%; Score 44; DB 8; Length 732;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 TKTFFSVKXDLISLPEQ 17
Db 374 TTTPTSTEDVISLDDQ 389

RESULT 40
ADQ09850
ID ADQ09850 standard; protein; 87 AA.

AC ADQ09850;

XX 23-SEP-2004 (first entry)

DE Mouse NK-2 class homeobox protein, NKx2.2.

XX Mouse; islet cell differentiation transcription factor;
KW insulin-dependent diabetes; insulin; somatic cell;
KW insulin-producing cell; islet cell; NeuroD; ngn3; neurogenin 3; Pax6;
KW paired-box transcription factor 6; Pax4;
KW paired-box transcription factor 4; Nkx2.2; NK-2 class homeobox protein;
KW Nkx6.1; Isl-1; Islet factor 1; Pdx-1;
KW pancreatic and duodenal homeobox protein; BCT; betacellulin.
XX Mus musculus.

OS US2004132679-A1.
XX 08-JUL-2004.

XX 03-SEP-2003; 2003US-00654102.

XX 03-SEP-2002; 2002US-0407743P.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX Chan L, Kojima H;

XX WPI; 2004-517032/49.

DR N-ESDB; ADQ09869, ADQ09943.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2005, 17:03:40 ; Search time 5.98311 Seconds
(without alignments)
286.963 Million cell updates/sec

Title: US-09-998-861-7
Perfect score: 102
Sequence: 1 TKTPFSVKDILSLPEQXRAXGA 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pdp:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pdp:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pdp:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pdp:*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pdp:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pdp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	66.7	324	4	US-09-949-016-6488
2	68	66.7	382	4	US-09-949-016-10839
3	60	58.8	86	3	US-09-008-892-4
4	60	58.8	273	3	US-09-008-892-2
5	50	49.0	344	4	US-09-489-039A-11932
6	46.5	45.6	371	2	US-08-442-809A-76
7	45	44.1	338	4	US-09-543-681A-7061
8	44	43.1	113	4	US-09-270-767-60627
9	44	43.1	296	4	US-09-270-767-45136
10	44	43.1	553	2	US-08-586-272-2
11	44	43.1	553	3	US-09-082-969-2
12	43	42.2	731	4	US-09-252-991A-22524
13	42.5	41.7	254	4	US-09-489-039A-12987
14	42	41.2	251	5	PCT-US96-01314-59
15	42	41.2	785	4	US-09-902-540-10007
16	42	41.2	798	2	US-07-728-215-30
17	42	41.2	798	3	US-08-938-085A-30
18	42	41.2	798	4	US-10-072-844-30
19	42	41.2	798	4	US-10-072-844-30
20	42	41.2	798	4	US-10-072-841A-30
21	42	41.2	798	4	US-10-219-631A-30
22	42	41.2	798	4	US-09-949-016-6193
23	42	41.2	968	4	US-09-107-532A-5970
24	41.5	40.7	454	4	US-09-902-540-12747
25	41	40.2	209	4	US-09-388-221B-19
26	41	40.2	216	4	US-09-902-540-11806
27	41	40.2	223	4	US-09-009-816-4

28	41	40.2	367	4	US-09-009-816-2	Sequence 2, Appli
29	41	40.2	1399	4	US-09-388-221B-4	Sequence 4, Appli
30	41	40.2	1424	4	US-09-388-221B-12	Sequence 12, Appli
31	41	40.2	1443	4	US-09-388-221B-6	Sequence 6, Appli
32	41	40.2	1454	4	US-09-388-221B-10	Sequence 10, Appli
33	41	40.2	1473	4	US-09-388-221B-2	Sequence 2, Appli
34	40.5	39.7	1376	4	US-09-949-016-6289	Sequence 6289, Ap
35	40	39.2	164	4	US-09-248-796A-18291	Sequence 18291, A
36	40	39.2	216	4	US-09-248-796A-18200	Sequence 18200, A
37	40	39.2	246	4	US-09-631-594-63	Sequence 63, Appli
38	40	39.2	264	4	US-09-631-594-64	Sequence 64, Appli
39	40	39.2	364	4	US-09-949-016-7730	Sequence 7730, Ap
40	40	39.2	364	4	US-09-949-016-7731	Sequence 7731, Ap
41	40	39.2	364	4	US-09-949-016-7732	Sequence 7732, Ap
42	40	39.2	364	4	US-09-949-016-7733	Sequence 7733, Ap
43	40	39.2	735	2	US-08-313-185-48	Sequence 48, Appli
44	40	39.2	735	2	US-08-459-499-9	Sequence 9, Appli
45	40	39.2	735	2	US-08-459-499-12	Sequence 12, Appli

ALIGNMENTS

RESULT 1

US-09-949-016-6488
; Sequence 6488, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6488
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6488

Query Match 66.7%; Score 68; DB 4; Length 324;
Best Local Similarity 63.6%; Pred. No. 0.0012;
Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Oy	2	TKTPFSVKDILSLPEQXRAXGA	23
Db	8	TPTPFSVKDILSLPEQXRAXGA	29

RESULT 2

US-09-949-016-10839
; Sequence 10839, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498

[illegible]

;; PRIOR FILING DATE: 1999-01-29
;; NUMBER OF SEQ ID NOS: 14342
;; SEQ ID NO 11932
;; LENGTH: 344
;; TYPE: PRT
;; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11932

Query Match 49.0%; Score 50; DB 4; Length 344;
Best Local Similarity 57.9%; Pred. No. 1.6;
Matches 11; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 TPFSVKDILSLPEQXRAG 22
|||||:|:|:|
Db 87 TPFSVETLSIETARHG 105

RESULT 6

US-08-442-809A-76
; Sequence 76, Application US/08442809A
; Patent No. 5976873

GENERAL INFORMATION:

;; APPLICANT: Bohinski, Robert J.,
;; APPLICANT: Whitsett, Jeffrey A.
;; TITLE OF INVENTION: Nucleic Acid Sequences
;; TITLE OF INVENTION: Controlling Lung Cell -
;; TITLE OF INVENTION: Specific Gene Expression
;; NUMBER OF SEQUENCES: 76
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
;; ADDRESSEE: Cecchi, Stewart & Olstein
;; STREET: 6 Becker Farm Road
;; CITY: Roseland
;; STATE: New Jersey
;; COUNTRY: USA
;; ZIP: 07068

COMPUTER READABLE FORM:

;; MEDIUM TYPE: 3.5 inch diskette
;; COMPUTER: IBM PS/2
;; OPERATING SYSTEM: MS-DOS
;; SOFTWARE: WordPerfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/442,809A
;; FILING DATE: 17-MAY-1995
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/245,356
;; FILING DATE: 18-MAY-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Olstein, Elliot M.
;; REGISTRATION NUMBER: 24,025
;; REFERENCE/DOCKET NUMBER: 271010-360
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 201-994-1700
;; TELEFAX: 201-994-1744
;; INFORMATION FOR SEQ ID NO: 76:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 371 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; FEATURE:
;; NAME/KEY: human thyroid transcription factor-1
US-08-442-809A-76

Query Match 45.6%; Score 46.5; DB 2; Length 371;
Best Local Similarity 60.0%; Pred. No. 6.9;
Matches 12; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 4 TPFSVKDILSLPEQXRAG 22
|||||:|:|:|
Db 9 TPFSVSDILSPLSEYKVG 28

RESULT 7

US-09-543-681A-7061
; Sequence 7061, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:

;; APPLICANT: GARY BRETON
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
;; FILE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
;; CURRENT APPLICATION NUMBER: US/09/543,681A
;; CURRENT FILING DATE: 2000-04-05
;; PRIOR APPLICATION NUMBER: US 60/128,706
;; PRIOR FILING DATE: 1999-04-09
;; NUMBER OF SEQ ID NOS: 8344
;; SEQ ID NO 7061
;; LENGTH: 338
;; TYPE: PRT
;; ORGANISM: Proteus mirabilis
US-09-543-681A-7061

Query Match 44.1%; Score 45; DB 4; Length 338;
Best Local Similarity 52.6%; Pred. No. 11;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 4 TPFSVKDILSLPEQXRAG 22
|||||:|:|:|
Db 81 TPFSVEMILSIERTARQFG 99

RESULT 8

US-09-270-767-60627
; Sequence 60627, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:

;; APPLICANT: Homburger et al.
;; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;; FILE REFERENCE: File Reference: 7326-094
;; CURRENT APPLICATION NUMBER: US/09/270,767
;; CURRENT FILING DATE: 1999-03-17
;; NUMBER OF SEQ ID NOS: 62517
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 60627
;; LENGTH: 113
;; TYPE: PRT
;; ORGANISM: Drosophila melanogaster
US-09-270-767-60627

Query Match 43.1%; Score 44; DB 4; Length 113;
Best Local Similarity 71.4%; Pred. No. 4.7;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 TPFSVKDILSLPEQ 17
|||||:|:|:|
Db 46 TPFSVTDILSPIEE 59

RESULT 9

US-09-270-767-45136
; Sequence 45136, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:

;; APPLICANT: Homburger et al.
;; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;; FILE REFERENCE: File Reference: 7326-094
;; CURRENT APPLICATION NUMBER: US/09/270,767
;; CURRENT FILING DATE: 1999-03-17
;; NUMBER OF SEQ ID NOS: 62517
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 45136
;; LENGTH: 296
;; TYPE: PRT

us-09-998-861-7.rai

Thu Apr 14 08:26:23 2005

```

; ORGANISM: Drosophila melanogaster
US-09-270-767-45136

Query Match 43.1%; Score 44; DB 4; Length 296;
Best Local Similarity 71.4%; Pred. No. 14;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 TPFVSKDILSLPEQ 17
DB 46 TPFVTDILSLPIEE 59

RESULT 10
US-08-586-272-2
; Sequence 2, Application US/08586272
; Patent No. 5824795
; GENERAL INFORMATION:
; APPLICANT: POPOFF, MICHEL Y.
; APPLICANT: LE GUERN FELLOUS, MURIEL
; TITLE OF INVENTION: OLIGONUCLEOTIDES FOR THE DETECTION OF
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESS: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/586,272
; FILING DATE: 16-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9500410
; FILING DATE: 16-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-107-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-586-272-2

Query Match 43.1%; Score 44; DB 2; Length 553;
Best Local Similarity 61.1%; Pred. No. 29;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 TKTFFSVKDILSLPEQXR 19
DB 164 TKNCRSVKDILELMDQLR 181

RESULT 11
US-09-082-969-2
; Sequence 2, Application US/09082969
; Patent No. 6080545
; GENERAL INFORMATION:
; APPLICANT: POPOFF, MICHEL Y.
; APPLICANT: LE GUERN FELLOUS, MURIEL
; TITLE OF INVENTION: OLIGONUCLEOTIDES FOR THE DETECTION OF
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESS: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/586,272
; FILING DATE: 16-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9500410
; FILING DATE: 16-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-107-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-586-272-2

Query Match 43.1%; Score 44; DB 2; Length 553;
Best Local Similarity 61.1%; Pred. No. 29;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 TKTFFSVKDILSLPEQXR 19
DB 164 TKNCRSVKDILELMDQLR 181

RESULT 12
US-09-252-991A-22524
; Sequence 22524, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22524
; LENGTH: 731
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22524

Query Match 42.2%; Score 43; DB 4; Length 731;
Best Local Similarity 61.5%; Pred. No. 60;

```



```
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 3 KTFPSVKDILSLP 15
DB 574 RTPASVDDILTP 586

RESULT 13
US-09-489-039A-12987
; Sequence 12987, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Bretton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12987
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12987

Query Match 41.7%; Score 42.5; DB 4; Length 254;
Best Local Similarity 47.8%; Pred. No. 22;
Matches 11; Conservative 3; Mismatches 6; Indels 3; Gaps 1;

QY 3 KTFPSVKDILSLPE---QVRXG 22
DB 171 RTPYTVTDIASLMKIAQVRROQ 193

RESULT 14
PCT-US96-01314-59
; Sequence 59, Application PC/TUS9601314
; GENERAL INFORMATION:
; APPLICANT: M. Amin Arnaut
; TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN
; TITLE OF INVENTION: ANTAGONISTS
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/01314
; FILING DATE: 30-JAN-96
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/380,167
; FILING DATE: 30-JAN-95
; ATTORNEY/AGENT INFORMATION:
; NAME: John W. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00786/267001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 251 amino acids
```

```
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-01314-59

Query Match 41.2%; Score 42; DB 5; Length 251;
Best Local Similarity 72.7%; Pred. No. 26;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 TPFPSVKDILSL 14
DB 90 TPFPSYKNVLSL 100

RESULT 15
US-09-902-540-10007
; Sequence 10007, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 10007
; LENGTH: 785
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-10007

Query Match 41.2%; Score 42; DB 4; Length 785;
Best Local Similarity 64.3%; Pred. No. 97;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 7 SVKDILSLPEQXRA 20
DB 660 SVQDITELPENLRA 673

RESULT 16
US-07-728-215-30
; Sequence 30, Application US/07728215
; Patent No. 5962643
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Dean
; APPLICANT: Quaranta, Vito
; APPLICANT: Pytela, Robert
; TITLE OF INVENTION: A No. 5962643el Integrin Beta Subunit and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: pretty, Schroeder, Brueggemann & Clark
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States of America
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/728,215
; FILING DATE: 19910711
; CLASSIFICATION: 435
```


ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P31 8717
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 798 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-728-215-30

Query Match 41.2%; Score 42; DB 2; Length 798;
Best Local Similarity 72.7%; Pred. No. 99;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 TPFSVKDILSL 14
||| |::|||
Db 215 TPFSYKNVLSL 225

RESULT 17
US-08-938-085A-30
Sequence 30, Application US/08938085A
Patent No. 6339148
GENERAL INFORMATION:
APPLICANT: Sheppard, Dean
APPLICANT: Quaranta, Vito
APPLICANT: Pytela, Robert
TITLE OF INVENTION: A No. 6339148el Integrin Beta Subunit and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,085A
FILING DATE: 26-SEP-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/728,215
FILING DATE: 11-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Parent, Annette S.
REGISTRATION NUMBER: 42,058
REFERENCE/DOCKET NUMBER: 023070-080210US
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 798 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-10-072-844-30

Query Match 41.2%; Score 42; DB 3; Length 798;
Best Local Similarity 72.7%; Pred. No. 99;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 TPFSVKDILSL 14
||| |::|||
Db 215 TPFSYKNVLSL 225

RESULT 18
US-10-072-844-30
Sequence 30, Application US/10072844
Patent No. 6576432
GENERAL INFORMATION:
APPLICANT: Sheppard, Dean
APPLICANT: Quaranta, Vito
APPLICANT: Pytela, Robert
TITLE OF INVENTION: A No. 6576432el Integrin Beta Subunit and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/072,844
FILING DATE: 06-Feb-2002
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/938,085A
FILING DATE: 26-SEP-1997
FILING DATE: 11-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Parent, Annette S.
REGISTRATION NUMBER: 42,058
REFERENCE/DOCKET NUMBER: 023070-080210US
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 798 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-10-072-844-30

Query Match 41.2%; Score 42; DB 4; Length 798;
Best Local Similarity 72.7%; Pred. No. 99;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 TPFSVKDILSL 14
||| |::|||
Db 215 TPFSYKNVLSL 225

RESULT 19
US-10-072-838-30
Sequence 30, Application US/10072838
Patent No. 6596277
GENERAL INFORMATION:
APPLICANT: Sheppard, Dean
APPLICANT: Quaranta, Vito
APPLICANT: Pytela, Robert
TITLE OF INVENTION: A No. 6596277el Integrin Beta Subunit and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/072,844
FILING DATE: 06-Feb-2002
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/728,215
FILING DATE: 11-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Parent, Annette S.
REGISTRATION NUMBER: 42,058
REFERENCE/DOCKET NUMBER: 023070-080210US
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 798 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-10-072-844-30

Query Match 41.2%; Score 42; DB 3; Length 798;
Best Local Similarity 72.7%; Pred. No. 99;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 TPFSVKDILSL 14
||| |::|||
Db 215 TPFSYKNVLSL 225

;/ Thereof
;/ NUMBER OF SEQUENCES: 43
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
;/ STREET: 4370 La Jolla Village Drive, Suite 700
;/ CITY: San Diego
;/ STATE: California
;/ COUNTRY: United States of America
;/ ZIP: 92122
;/
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: PatentIn Release #1.0, Version #1.25
;/
;/ CURRENT APPLICATION DATA: US/10/072,838
;/ APPLICATION NUMBER: US/10/072,838
;/ FILING DATE: 06-Feb-2002
;/ CLASSIFICATION: <Unknown>
;/
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: 07/728,215
;/ FILING DATE: <Unknown>
;/
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Campbell, Cathryn A.
;/ REGISTRATION NUMBER: 31,815
;/ REFERENCE/DOCKET NUMBER: P31 8717
;/
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (619) 535-9001
;/ TELEFAX: (619) 535-8949
;/
;/ INFORMATION FOR SEQ ID NO: 30:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 798 amino acids
;/ TYPE: amino acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/
;/ SEQUENCE DESCRIPTION: SEQ ID NO: 30:
;/ US-10-072-838-30
;/
;/ Query Match 41.2%; Score 42; DB 4; Length 798;
;/ Best Local Similarity 72.7%; Pred. No. 99;
;/ Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
;/
;/ QY 4 TPFSVKDILSL 14
;/ ||||| :|||
;/ Db 215 TPFSYKNVLSL 225
;/
;/ RESULT 20
;/ US-10-072-841A-30
;/ Sequence 30, Application US/10072841A
;/ Patent No. 6639056
;/ GENERAL INFORMATION:
;/ APPLICANT: Sheppard, Dean
;/ Pytela, Robert
;/
;/ TITLE OF INVENTION: A No. 6639056el Integrin Beta Subunit and Uses
;/
;/ NUMBER OF SEQUENCES: 62
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Townsend and Townsend and Crew LLP
;/ STREET: Two Embarcadero Center, Eighth Floor
;/ CITY: San Francisco
;/ STATE: California
;/ COUNTRY: USA
;/ ZIP: 94111-3834
;/
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: PatentIn Release #1.0, Version #1.30
;/
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/10/072,841A
;/ FILING DATE: 02-Jun-2002
;/ CLASSIFICATION: 435
;/
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/938,085
;/ FILING DATE: 26-SEP-1997
;/ APPLICATION NUMBER: US 07/728,215
;/ FILING DATE: 11-JUL-1991
;/
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Parent, Annette S.
;/ REGISTRATION NUMBER: 42,058
;/ REFERENCE/DOCKET NUMBER: 023070-080210US
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (415) 576-0200
;/ TELEFAX: (415) 576-0300
;/
;/ INFORMATION FOR SEQ ID NO: 30:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 798 amino acids

;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/938,085A
;/ FILING DATE: 26-SEP-1997
;/ APPLICATION NUMBER: US 07/728,215
;/ FILING DATE: 11-JUL-1991
;/
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Parent, Annette S.
;/ REGISTRATION NUMBER: 42,058
;/ REFERENCE/DOCKET NUMBER: 023070-080210US
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (415) 576-0200
;/ TELEFAX: (415) 576-0300
;/
;/ INFORMATION FOR SEQ ID NO: 30:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 798 amino acids
;/ TYPE: amino acid
;/ STRANDEDNESS: <Unknown>
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: protein
;/ SEQUENCE DESCRIPTION: SEQ ID NO: 30:
;/ US-10-072-841A-30
;/
;/ Query Match 41.2%; Score 42; DB 4; Length 798;
;/ Best Local Similarity 72.7%; Pred. No. 99;
;/ Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
;/
;/ QY 4 TPFSVKDILSL 14
;/ ||||| :|||
;/ Db 215 TPFSYKNVLSL 225
;/
;/ RESULT 21
;/ US-10-219-631A-30
;/ Sequence 30, Application US/10219631A
;/ Patent No. 6787322
;/ GENERAL INFORMATION:
;/ APPLICANT: Sheppard, Dean
;/ Quaranta, Vito
;/ Pytela, Robert
;/
;/ TITLE OF INVENTION: A No. 6787322el Integrin Beta Subunit and Uses
;/
;/ NUMBER OF SEQUENCES: 62
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Townsend and Townsend and Crew LLP
;/ STREET: Two Embarcadero Center, Eighth Floor
;/ CITY: San Francisco
;/ STATE: California
;/ COUNTRY: USA
;/ ZIP: 94111-3834
;/
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: PatentIn Release #1.0, Version #1.30
;/
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/10/219,631A
;/ FILING DATE: 14-Aug-2002
;/ CLASSIFICATION: 435
;/
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/938,085
;/ FILING DATE: 26-SEP-1997
;/ APPLICATION NUMBER: US 07/728,215
;/ FILING DATE: 11-JUL-1991
;/
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Parent, Annette S.
;/ REGISTRATION NUMBER: 42,058
;/ REFERENCE/DOCKET NUMBER: 023070-080210US
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (415) 576-0200
;/ TELEFAX: (415) 576-0300
;/
;/ INFORMATION FOR SEQ ID NO: 30:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 798 amino acids


```

; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-10-219-631A-30

Query Match          41.2%; Score 42; DB 4; Length 798;
Best Local Similarity 72.7%; Pred. No. 99;
Matches 8; Conservative 2; Mismatches 1; Indels 1; Gaps 0;

QY      4 TPFSVKDILSL 14
DB      215 TPFSYKNVLSL 225

RESULT 22
US-09-949-016-6193
; Sequence 6193, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6193
; LENGTH: 798
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6193

Query Match          41.2%; Score 42; DB 4; Length 798;
Best Local Similarity 72.7%; Pred. No. 99;
Matches 8; Conservative 2; Mismatches 1; Indels 1; Gaps 0;

QY      4 TPFSVKDILSL 14
DB      215 TPFSYKNVLSL 225

RESULT 23
US-09-107-532A-5970
; Sequence 5970, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A

```

```

; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5970:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 968 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...968
; SEQUENCE DESCRIPTION: SEQ ID NO: 5970:
US-09-107-532A-5970

Query Match          41.2%; Score 42; DB 4; Length 968;
Best Local Similarity 69.2%; Pred. No. 1.2e+02;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      4 TPFSVKDILSLPE 16
DB      787 TVFAVKDGVSLPE 799

RESULT 24
US-09-902-540-12747
; Sequence 12747, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 12747
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
; US-09-902-540-12747

Query Match          40.7%; Score 41.5; DB 4; Length 454;
Best Local Similarity 56.2%; Pred. No. 63;
Matches 9; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY      2 TKTFFSVKDILSLPEQ 17
DB      5 TSAPFDIPD-LSLPQQ 19

RESULT 25
US-09-388-221B-19
; Sequence 19, Application US/09388221B
; Patent No. 6818750
; GENERAL INFORMATION:

```


; APPLICANT: Reed, John C.
; TITLE OF INVENTION: No. 6818750el Card Proteins Involved in Cell Death Regulation
; FILE REFERENCE: P-LJ 3650
; CURRENT APPLICATION NUMBER: US/09/388,221B
; CURRENT FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-388-221B-19

Query Match 40.2%; Score 41; DB 4; Length 209;
Best Local Similarity 50.0%; Pred. No. 31;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 TKTPFSVKDILSLPEQ 17
||| : : ||| :
Db 61 TATPAPIRQLSRPER 76

RESULT 26
US-09-902-540-11806
; Sequence 11806, Application US/09502540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 11806
; LENGTH: 216
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-11806

Query Match 40.2%; Score 41; DB 4; Length 216;
Best Local Similarity 46.7%; Pred. No. 32;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 TKTPFSVKDILSLPE 16
||| : : ||| :
Db 22 TKRPATYEDLVALPE 36

RESULT 27
US-09-009-816-4
; Sequence 4, Application US/09009816
; Patent No. 6436667
; GENERAL INFORMATION:
; APPLICANT: German, Michael
; APPLICANT: Permutt, M. Alan
; APPLICANT: Inoue, Hiroshi
; TITLE OF INVENTION: Human Nkx-6.1 Polypeptide-Encoding
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,816
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Francis, Carol L
; REGISTRATION NUMBER: 36,513
; REFERENCE/DOCKET NUMBER: 9076/082CIP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3400
; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-009-816-4

Query Match 40.2%; Score 41; DB 4; Length 223;
Best Local Similarity 45.0%; Pred. No. 34;
Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 4 TPFSVKDILSLPEQXRAXGA 23
||| : : ||| :
Db 94 TPHGINNLSRPSMPVASGA 113

RESULT 28
US-09-009-816-2
; Sequence 2, Application US/09009816
; Patent No. 6436667
; GENERAL INFORMATION:
; APPLICANT: German, Michael
; APPLICANT: Permutt, M. Alan
; APPLICANT: Inoue, Hiroshi
; TITLE OF INVENTION: Human Nkx-6.1 Polypeptide-Encoding
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,816
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Francis, Carol L
; REGISTRATION NUMBER: 36,513
; REFERENCE/DOCKET NUMBER: 9076/082CIP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3400

TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 367 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-09-009-816-2

Query Match 40.2%; Score 41; DB 4; Length 367;
Best Local Similarity 45.0%; Pred. No. 60;
Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 4 TPFVSKDILSLPEQXRXGA 23
DB 94 TPGINNILSRPSPMPVASGA 113

RESULT 29
US-09-388-221B-4
; Sequence 4, Application US/09388221B
; Patent No. 6818750
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: No. 6818750el Card Proteins Involved in Cell Death Regulation
; FILE REFERENCE: P-LJ 3650
; CURRENT APPLICATION NUMBER: US/09/388,221B
; CURRENT FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1399
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-388-221B-4

Query Match 40.2%; Score 41; DB 4; Length 1399;
Best Local Similarity 50.0%; Pred. No. 2.8e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 TKTPFSVKDILSLPEQ 17
DB 389 TATPAPIRQILSRPER 404

RESULT 30
US-09-388-221B-12
; Sequence 12, Application US/09388221B
; Patent No. 6818750
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: No. 6818750el Card Proteins Involved in Cell Death Regulation
; FILE REFERENCE: P-LJ 3650
; CURRENT APPLICATION NUMBER: US/09/388,221B
; CURRENT FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 1424
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Construct
US-09-388-221B-12

Query Match 40.2%; Score 41; DB 4; Length 1424;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 TKTPFSVKDILSLPEQ 17
DB 389 TATPAPIRQILSRPER 404

RESULT 31
US-09-388-221B-6
; Sequence 6, Application US/09388221B
; Patent No. 6818750
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: No. 6818750el Card Proteins Involved in Cell Death Regulation
; FILE REFERENCE: P-LJ 3650
; CURRENT APPLICATION NUMBER: US/09/388,221B
; CURRENT FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1443
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-388-221B-6

Query Match 40.2%; Score 41; DB 4; Length 1443;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 TKTPFSVKDILSLPEQ 17
DB 389 TATPAPIRQILSRPER 404

RESULT 32
US-09-388-221B-10
; Sequence 10, Application US/09388221B
; Patent No. 6818750
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: No. 6818750el Card Proteins Involved in Cell Death Regulation
; FILE REFERENCE: P-LJ 3650
; CURRENT APPLICATION NUMBER: US/09/388,221B
; CURRENT FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1454
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Construct
US-09-388-221B-10

Query Match 40.2%; Score 41; DB 4; Length 1454;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 TKTPFSVKDILSLPEQ 17
DB 389 TATPAPIRQILSRPER 404

RESULT 33
US-09-388-221B-2
; Sequence 2, Application US/09388221B
; Patent No. 6818750
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: No. 6818750el Card Proteins Involved in Cell Death Regulation
; FILE REFERENCE: P-LJ 3650
; CURRENT APPLICATION NUMBER: US/09/388,221B
; CURRENT FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2

TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 367 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-09-009-816-2

Query Match 40.2%; Score 41; DB 4; Length 367;
Best Local Similarity 45.0%; Pred. No. 60;
Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 4 TPFVSKDILSLPEQXRXGA 23
DB 94 TPGINNILSRPSPMPVASGA 113

RESULT 29
US-09-388-221B-4
; Sequence 4, Application US/09388221B
; Patent No. 6818750
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: No. 6818750el Card Proteins Involved in Cell Death Regulation
; FILE REFERENCE: P-LJ 3650
; CURRENT APPLICATION NUMBER: US/09/388,221B
; CURRENT FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1399
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-388-221B-4

Query Match 40.2%; Score 41; DB 4; Length 1399;
Best Local Similarity 50.0%; Pred. No. 2.8e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 TKTPFSVKDILSLPEQ 17
DB 389 TATPAPIRQILSRPER 404

RESULT 30
US-09-388-221B-12
; Sequence 12, Application US/09388221B
; Patent No. 6818750
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: No. 6818750el Card Proteins Involved in Cell Death Regulation
; FILE REFERENCE: P-LJ 3650
; CURRENT APPLICATION NUMBER: US/09/388,221B
; CURRENT FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 1424
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Construct
US-09-388-221B-12

Query Match 40.2%; Score 41; DB 4; Length 1424;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 TKTPFSVKDILSLPEQ 17
DB 389 TATPAPIRQILSRPER 404

RESULT 31
US-09-388-221B-6
; Sequence 6, Application US/09388221B
; Patent No. 6818750
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: No. 6818750el Card Proteins Involved in Cell Death Regulation
; FILE REFERENCE: P-LJ 3650
; CURRENT APPLICATION NUMBER: US/09/388,221B
; CURRENT FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1443
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-388-221B-6

Query Match 40.2%; Score 41; DB 4; Length 1443;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 TKTPFSVKDILSLPEQ 17
DB 389 TATPAPIRQILSRPER 404

RESULT 32
US-09-388-221B-10
; Sequence 10, Application US/09388221B
; Patent No. 6818750
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: No. 6818750el Card Proteins Involved in Cell Death Regulation
; FILE REFERENCE: P-LJ 3650
; CURRENT APPLICATION NUMBER: US/09/388,221B
; CURRENT FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1454
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Construct
US-09-388-221B-10

Query Match 40.2%; Score 41; DB 4; Length 1454;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 TKTPFSVKDILSLPEQ 17
DB 389 TATPAPIRQILSRPER 404

RESULT 33
US-09-388-221B-2
; Sequence 2, Application US/09388221B
; Patent No. 6818750
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: No. 6818750el Card Proteins Involved in Cell Death Regulation
; FILE REFERENCE: P-LJ 3650
; CURRENT APPLICATION NUMBER: US/09/388,221B
; CURRENT FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2

; LENGTH: 1473
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-388-221B-2

Query Match 40.2%; Score 41; DB 4; Length 1473;
Best Local Similarity 50.0%; Pred. No. 3e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 TTPFSVKDILSLPEQ 17
DB 389 TATPAPIRQLSRPER 404

RESULT 34
US-09-949-016-6289
; Sequence 6289, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6289
; LENGTH: 1376
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6289

Query Match 39.7%; Score 40.5; DB 4; Length 1376;
Best Local Similarity 54.5%; Pred. No. 3.4e+02;
Matches 12; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 3 KTFPSVKDILSLPEQXRAXGA 23
DB 979 KAGFSIKDILSLCERHNGINGA 1000

RESULT 35
US-09-248-796A-18291
; Sequence 18291, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196-132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18291
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-18291

Query Match 39.2%; Score 40; DB 4; Length 164;
Best Local Similarity 37.5%; Pred. No. 35;
Matches 6; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 4 TTPFSVKDILSLPEQXR 19
DB 91 SPFLLEVLNNPEQAK 106

RESULT 36
US-09-248-796A-18200
; Sequence 18200, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196-132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18200
; LENGTH: 216
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-18200

Query Match 39.2%; Score 40; DB 4; Length 216;
Best Local Similarity 57.1%; Pred. No. 48;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 KTFPSVKDILSLPE 16
DB 154 KTKRPMKDVGLPE 167

RESULT 37
US-09-631-594-63
; Sequence 63, Application US/09631594
; Patent No. 6737237
; GENERAL INFORMATION:
; APPLICANT: MCLEOD, RINA W.
; APPLICANT: ROBERTS, CRAIG W.
; APPLICANT: ROBERTS, FIONA
; APPLICANT: JOHNSON, JENNIFER J.
; APPLICANT: KIRISITS, MICHAEL
; APPLICANT: FERGUSON, DAVID
; APPLICANT: LYONS, RUSSELL
; APPLICANT: MOI, ERNEST
; APPLICANT: MACK, DOUG
; APPLICANT: SAMUEL, BENJAMIN
; APPLICANT: GORNICKI, PIOTR
; APPLICANT: ZUTHER, ELLEN
; TITLE OF INVENTION: ANTIMICROBIAL AGENTS, DIAGNOSTIC REAGENTS, AND VACCINES
; TITLE OF INVENTION: BASED ON UNIQUE APICOMPLEXAN PARASITE COMPONENTS
; FILE REFERENCE: 19338-90966
; CURRENT APPLICATION NUMBER: US/09/631,594
; CURRENT FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: PCT/US00/11478
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US97/12497
; PRIOR FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Glycine max
US-09-631-594-63

Query Match 39.2%; Score 40; DB 4; Length 246;
Best Local Similarity 46.2%; Pred. No. 56;

Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TKTFSVKDILSL 14
|:|:|:|:|:
Db 19 TRPYTARDVVS 31

RESULT 39
US-09-631-594-64
; Sequence 64, Application US/09631594
; Patent No. 6737237
; GENERAL INFORMATION:
; APPLICANT: MCLEOD, RIMA W.
; APPLICANT: ROBERTS, CRAIG W.
; APPLICANT: ROBERTS, JENNA
; APPLICANT: JOHNSON, JENNIFER J.
; APPLICANT: KIRISITS, MICHAEL
; APPLICANT: FERGUSON, DAVID
; APPLICANT: LYONS, RUSSELL
; APPLICANT: MUL, ERNEST
; APPLICANT: MACK, DOUG
; APPLICANT: SAMUEL, BENJAMIN
; APPLICANT: GORNICKI, PIOTR
; APPLICANT: ZUTHER, ELLEN
; TITLE OF INVENTION: ANTIMICROBIAL AGENTS, DIAGNOSTIC REAGENTS, AND VACCINES
; FILE REFERENCE: 19338-90966
; CURRENT APPLICATION NUMBER: US/09/631,594
; CURRENT FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: PCT/US00/11478
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US97/12497
; PRIOR FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 64
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Cucurbita sp.
US-09-631-594-64

Query Match 39.2%; Score 40; DB 4; Length 264;
Best Local Similarity 46.2%; Pred. No. 61;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TKTFSVKDILSL 14
|:|:|:|:|:
Db 37 TRPYTAKDVVSL 49

RESULT 39
US-09-949-016-7730
; Sequence 7730, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7730
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Human

US-09-949-016-7730

Query Match 39.2%; Score 40; DB 4; Length 364;
Best Local Similarity 52.6%; Pred. No. 88;
Matches 10; Conservative 3; Mismatches 4; Indels 2; Gaps 1;

Qy 4 TPFSVKDIL--SLPEQXRA 20
|:|:|:|:|:
Db 256 SPWSVPSVLWTSCEQNRA 274

RESULT 40
US-09-949-016-7731
; Sequence 7731, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7731
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7731

Query Match 39.2%; Score 40; DB 4; Length 364;
Best Local Similarity 52.6%; Pred. No. 88;
Matches 10; Conservative 3; Mismatches 4; Indels 2; Gaps 1;

Qy 4 TPFSVKDIL--SLPEQXRA 20
|:|:|:|:|:
Db 256 SPWSVPSVLWTSCEQNRA.274

Search completed: April 13, 2005, 17:18:48
Job time : 6.98311 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2005, 17:06:25 ; Search time 198.27 Seconds

(without alignments)
439.204 Million cell updates/sec

Title: US-09-998-861-13

Perfect score: 1396

Sequence: 1 MESNLQGTFLNNNTLAQFS.....LLLRKRAAFSVLSLGAHSV 262

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1421835 seqs, 332370683 residues

Total number of hits satisfying chosen parameters: 1421835

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1385	99.2	262	15	US-10-362-437-7
2	612	43.8	272	15	US-10-362-437-10
3	590	42.3	277	15	US-10-362-437-6
4	583	41.8	277	15	US-10-362-437-9
5	558.5	40.0	365	16	US-10-654-102-90
6	557	39.9	367	15	US-10-362-437-1
7	557	39.9	367	16	US-10-654-102-89
8	549.5	39.4	365	15	US-10-362-437-8
9	217	15.5	284	16	US-10-654-102-56
10	217	15.5	284	16	US-10-654-102-61
11	217	15.5	284	16	US-10-654-102-63
12	217	15.5	284	16	US-10-654-102-64
13	217	15.5	284	16	US-10-654-102-65

14	217	15.5	284	17	US-10-877-706-2	Sequence 2, Appli
15	217	15.5	284	17	US-10-475-021A-2	Sequence 2, Appli
16	212.5	15.2	328	9	US-09-900-527-2	Sequence 2, Appli
17	212.5	15.2	328	14	US-10-157-031-12	Sequence 12, Appli
18	211	15.1	284	9	US-09-759-847-2	Sequence 2, Appli
19	208	14.9	300	10	US-09-820-598-1	Sequence 1, Appli
20	208	14.9	300	13	US-10-095-932-1	Sequence 1, Appli
21	206.5	14.8	217	14	US-10-097-340-129	Sequence 129, App
22	205.5	14.7	217	15	US-10-097-105-1561	Sequence 1561, Ap
23	205.5	14.7	373	16	US-10-408-765A-975	Sequence 975, App
24	202	14.5	283	16	US-10-654-102-59	Sequence 59, Appl
25	202	14.5	283	16	US-10-654-102-67	Sequence 67, Appl
26	200.5	14.4	349	10	US-09-820-598-3	Sequence 3, Appli
27	200.5	14.4	349	13	US-10-095-932-3	Sequence 3, Appli
28	195	14.0	283	16	US-10-654-102-58	Sequence 58, Appl
29	195	14.0	283	16	US-10-654-102-60	Sequence 60, Appl
30	191	13.7	289	14	US-10-097-340-71	Sequence 71, Appl
31	191	13.7	289	15	US-10-295-027-626	Sequence 626, App
32	190.5	13.6	252	15	US-10-094-749-1844	Sequence 1844, Ap
33	188	13.5	311	14	US-10-012-456A-38	Sequence 38, Appl
34	187.5	13.4	254	14	US-10-012-456A-54	Sequence 54, Appl
35	186	13.3	312	16	US-10-322-281-672	Sequence 672, App
36	186	13.3	340	16	US-10-322-281-669	Sequence 669, App
37	185	13.3	149	16	US-10-755-889-552	Sequence 552, App
38	185	13.3	309	15	US-10-408-501-6	Sequence 6, Appli
39	185	13.3	314	15	US-10-408-501-2	Sequence 2, Appli
40	185	13.3	314	17	US-10-891-585-2	Sequence 2, Appli
41	185	13.3	314	17	US-10-891-585-3	Sequence 3, Appli
42	184.5	13.2	255	15	US-10-291-172-345	Sequence 345, App
43	184.5	13.2	255	15	US-10-221-278-345	Sequence 345, App
44	184.5	13.2	301	9	US-09-850-258-6	Sequence 6, Appli
45	184.5	13.2	304	9	US-09-850-258-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1

US-10-362-437-7
; Sequence 7, Application US/10362437
; Publication No. US20040053210A1
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M.
; APPLICANT: Briscoe, James
; APPLICANT: Ericson, Johan
; APPLICANT: Rubenstein, John L.
; APPLICANT: Sander, Maïke
; TITLE OF INVENTION: GENETIC DEMONSTRATION OF REQUIREMENT FOR NCK6.1, NCK2.2 AND NCK6
; FILE REFERENCE: 62166-A-PCT-US
; CURRENT APPLICATION NUMBER: US/10/362,437
; CURRENT FILING DATE: 2003-08-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Mouse
US-10-362-437-7

Query Match	99.2%	Score	1385;	DB	15;	Length	262;
Best Local Similarity	99.2%	Pred. No.	7.2e-126;				
Matches	260;	Conservative	1;	Mismatches	1;	Indels	0;
Qy	1	MESNLQGTFLNNNTLAQFS	EMKAPMCOYSQVNSFYKLSPPCLGQPLAAGT	PHG	ITD	ILS	60
Db	1	MESNLQGTFLNNNTLAQFS	EMKAPMCOYSQVNSFYKLSPPCLGQPLAAGT	PHG	ITD	ILS	60
Qy	61	RPVATPNSLLSGYPHVAGFGLSSQGVYVQGVQFSKAGNEY	PTRTNCWADT	QCDWR			120
Db	61	RPVATPNSLLSGYPHVAGFGLSSQGVYVQGVQFSKAGNEY	PTRTNCWADT	QCDWR			120
Qy	121	GSARPCGNTPDPLSTIHKKHTRPTFTTGHQI	FALEKTFEOTKYLAGPERAR	LAYS	LGMT		180


```

121 GSARPCSNTPDPLSDTIHKKHTRPTTGHQIFALEKTFEOTKYLAGPERARLAYSIGMT 180
181 ESQVKVWFONRTKWKSALEPSSSTPRAPGASGDRAASENEDDEVNKLDPDSDDEK 240
181 ESQVKVWFONRTKWKSALEPSSSTPRAPGASGDRAASENEDDEVNKLDPDSDNEK 240
241 IRLLRKHRAAFSVLSGAHSV 262
241 IRLLRKHRAAFSVLSGAHSV 262

RESULT 2
US-10-362-437-10
; Sequence 10, Application US/10362437
; Publication No. US20040053210A1
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M.
; APPLICANT: Briscoe, James
; APPLICANT: Ericson, Johan
; APPLICANT: Rubenstein, John L.
; APPLICANT: Sander, Maïke
; TITLE OF INVENTION: GENETIC DEMONSTRATION OF REQUIREMENT FOR NKX6.1, NKX2.2 AND NKX6.
; FILE OF INVENTION: VENTRAL NEURON GENERATION
; FILE REFERENCE: 62166-A-PCT-US
; CURRENT APPLICATION NUMBER: US/10/362.437
; CURRENT FILING DATE: 2003-08-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Chick
US-10-362-437-10

Query Match 43.8%; Score 612; DB 15; Length 272;
Best Local Similarity 52.5%; Pred. No. 8.2e-51;
Matches 136; Conservative 30; Mismatches 81; Indels 12; Gaps 7;

Qy 1 MESLQGTFLNNTOLA--QFSEMKAPMCOYSVQN--SFYKLSPPGIGLPOLAAGTPHGI 55
Db 1 MDNRQSAFVLGSTPLAALHNMAEMKSLFPYALQNSPFFKAPALGGLNLTQPLGTPHGI 60

Qy 56 TDILSRPVATPNSLLSGYPHVAGFGLSSQGVYVGPVQV--SFSKAGNEYPTRTNCW 114
Db 61 SDILGRPVCAA--GNLGLGPRINGLA--ASAGVYFGPAASRYRKPPLAELPGRPIFW 117

Qy 115 T--GODWRGSARPCGNTDPLSDTIHKKHTRPTTGHQIFALEKTFEOTKYLAGPERAR 172
Db 118 VVQGSFWRDPRITCPAQTGMVLDKDKKKHRSPTFSGQOIFALEKTFEOTKYLAGPERAR 177

Qy 173 LAYSIGMTESQVKVWFONRTKWKSALEPSSSTPRAPGASGDRAASENEDDEVNKL 231
Db 178 LAYSIGMTESQVKVWFONRTKWKSALEPSSSTPRAPGASGDRAASENEDDEVNKL 237

Qy 232 LDPDSDDEKIRLLLRKHRA 250
Db 238 LDPDSDDEKIRLLLRKHKS 256

RESULT 3
US-10-362-437-6
; Sequence 6, Application US/10362437
; Publication No. US20040053210A1
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M.
; APPLICANT: Briscoe, James
; APPLICANT: Ericson, Johan
; APPLICANT: Rubenstein, John L.
; APPLICANT: Sander, Maïke
; TITLE OF INVENTION: GENETIC DEMONSTRATION OF REQUIREMENT FOR NKX6.1, NKX2.2 AND NKX6.
; FILE REFERENCE: 62166-A-PCT-US

```

```

; CURRENT APPLICATION NUMBER: US/10/362.437
; CURRENT FILING DATE: 2003-08-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 6
; LENGTH: 277
; TYPE: PRT
; ORGANISM: Human
US-10-362-437-6

Query Match 42.3%; Score 590; DB 15; Length 277;
Best Local Similarity 50.8%; Pred. No. 1.2e-48;
Matches 133; Conservative 28; Mismatches 85; Indels 16; Gaps 8;

Qy 1 MESLQGTFLNNTOLA--QFSEMKAPMCOYSVQN--SFYKLSPPGIGLPOLAAGTPHGI 55
Db 1 MDNRPGAFVLSSAPLAALHNMAEMKSLFPYALQNSPFFKAPALGGLGQAQLPLGTPHGI 60

Qy 56 TDILSRPVATPNSLLSGYPHVAGFGLSSQGVYVGPVQV--SFSKAGNEYPTRTNCW 112
Db 61 SDILGRPVGAAGGLGGLPRLINGLA--SSAGVYFGPAASRYRKPPLAELPGRPIFW 118

Qy 113 ADT--GODWRGSARPCGNTDPLSDTIHKKHTRPTTGHQIFALEKTFEOTKYLAGPE 169
Db 119 PGVVOGAPWR--DRLAGPAGAGVLDKDKKKHRSPTFSGQOIFALEKTFEOTKYLAGPE 177

Qy 170 RARLAYSIGMTESQVKVWFONRTKWKSALEPSSSTPRAPGASGDRAASENEDDEVN 227
Db 178 RARLAYSIGMTESQVKVWFONRTKWKSALEPSSSTPRAPGASGDRAASENEDDEVN 237

Qy 228 YNKPLDPDSDDEKIRLLLRKH 249
Db 238 YNRPLDPDSDDEKIRLLLRKH 259

RESULT 4
US-10-362-437-9
; Sequence 9, Application US/10362437
; Publication No. US20040053210A1
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M.
; APPLICANT: Briscoe, James
; APPLICANT: Ericson, Johan
; APPLICANT: Rubenstein, John L.
; APPLICANT: Sander, Maïke
; TITLE OF INVENTION: GENETIC DEMONSTRATION OF REQUIREMENT FOR NKX6.1, NKX2.2 AND NKX6.
; FILE REFERENCE: 62166-A-PCT-US
; CURRENT APPLICATION NUMBER: US/10/362.437
; CURRENT FILING DATE: 2003-08-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 9
; LENGTH: 277
; TYPE: PRT
; ORGANISM: Mouse
US-10-362-437-9

Query Match 41.8%; Score 583; DB 15; Length 277;
Best Local Similarity 49.6%; Pred. No. 5.5e-48;
Matches 132; Conservative 29; Mismatches 81; Indels 24; Gaps 8;

Qy 1 MESLQGTFLNNTOLA--QFSEMKAPMCOYSVQN--SFYKLSPPGIGLPOLAAGTPHGI 55
Db 1 MDNRPGAFVLSSAPLAALHNMAEMKSLFPYALQNSPFFKAPALGGLGQAQLPLGTPHGI 60

Qy 56 TDILSRPVATPNSLLSGYPHVAGFGLSSQGVYVGPVQV--SFSKAGNEYPTRTNCW 112
Db 61 SDILGRPVGAAGGLGGLPRLINGLA--SSAGVYFGPAASRYRKPPLAELPGRPIFW 118

Qy 113 ADT--GODWR-----GSARPCGNTDPLSDTIHKKHTRPTTGHQIFALEKTFEOTKY 165
Db 119 PGVVOGSPWRDPRLAGSAQAAGG-----VLDKDKKKHRSPTFSGQOIFALEKTFEOTKY 173

```



```

Db 184 VAAVGRYKPLAELFGRTPIFWPGVMQSPWRDLAC--TPHQSILLDXDGRKXTRP 241
Qy 146 TFGTHQIFALEKTEQTKYLAGEPARARLAYSLGTMESQVQVWFQNRRTKWKKSALBESS 205
Db 242 TFGSQQIFALEKTEQTKYLAGEPARARLAYSLGTMESQVQVWFQNRRTKWKKHAEMAT 301
Qy 206 STPRAPGASGDRAASEN--EDDEYNKPLDPDSDDEKIRLLLRKHRAA 251
Db 302 AKKQDSETERLKGAASENEEDDDYDKPLDPSNDDKITQLLKKHKSS 349

RESULT 8
US-10-362-437-8
; Sequence 8, Application US/10362437
; Publication No. US20040053210A1
; GENERAL INFORMATION:
; APPLICANT: Jesseell, Thomas M.
; APPLICANT: Briscoe, James
; APPLICANT: Ericson, Johan
; APPLICANT: Rubenstein, John L.
; APPLICANT: Sander, Maïke
; TITLE OF INVENTION: GENETIC DEMONSTRATION OF REQUIREMENT FOR NKX6.1, NKX2.2 AND NKX6.
; TITLE OF INVENTION: VENTRAL NEURON GENERATION
; FILE REFERENCE: 62166-A-PCT-US
; CURRENT FILING DATE: 2003-08-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Rat
US-10-362-437-8

Query Match 39.4%; Score 549.5; DB 15; Length 365;
Best Local Similarity 41.2%; Pred. No. 1.4e-44;
Matches 147; Conservative 26; Mismatches 81; Indels 103; Gaps 13;

Qy 1 MESNLOGTFLANNITOLA--QFSEMKAPM-----CQYSVONSFYKLSPP-- 41
Db 7 MEGPRQSAFLUSSPPPLAALHSMWAEKMTPLYPAAYPPLTGPFPSSSSSSSSSPPLGA 66
Qy 42 ----GLGP-----QLAAGTGHGTTILSRP-----VATPNS-- 69
Db 67 HNPGLKPPAAGLSSLPQQLSAAATPHGINDILSRPMSFVAGSAGALPSASPSGSSS 126
Qy 70 -----LLSGYPHVAGGGLS-----SOGVYVGPQ-- 93
Db 127 SSSASATSAASAAAAAASAPAGLAGLPR---FSSLSPFPFPGLYFSPSAA 183
Qy 94 ----VGSFSKAGNEYPTRNCAWDTGQD--WRGSARPCGNTDP--LSDTIHKKHTR 144
Db 184 AVAAGRYPKPLAELFGRTPIFWPGVMQSPWRDLAC--TPHQSILLDXDGRKXTR 241
Qy 145 PTFTGHQIFALEKTEQTKYLAGEPARARLAYSLGTMESQVQVWFQNRRTKWKKSALBESS 204
Db 242 PTFSQQIFALEKTEQTKYLAGEPARARLAYSLGTMESQVQVWFQNRRTKWKKHAEMA 301
Qy 205 SSTPRAPGASGDRAASEN--DDYKNKPLDPDSDDEKIRLLLRKHRAAFSVLSIGA 259
Db 302 TAKKQDSETERLKGTSENEEDDDYDKPLDPSNDDKITQLLKKHKSSGGLLIHA 358

RESULT 9
US-10-654-102-56
; Sequence 56, Application US/10654102
; Publication No. US20040132679A1
; GENERAL INFORMATION:
; APPLICANT: CHAN, LAWRENCE
; APPLICANT: KOJIMA, HIDETO
; TITLE OF INVENTION: INDUCTION OF PANCREATIC ISLET FORMATION
; FILE REFERENCE: P024090S1

```

```

; CURRENT APPLICATION NUMBER: US/10/654,102
; CURRENT FILING DATE: 2003-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 56
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-654-102-56

Query Match 15.5%; Score 217; DB 16; Length 284;
Best Local Similarity 30.3%; Pred. No. 1.9e-12;
Matches 70; Conservative 26; Mismatches 95; Indels 40; Gaps 9;

Qy 25 PMCOYSVQNSFYKLSPPGLGPQ-----LAAAGTGHGTTILSRPVATPNSLLSGYPHV 77
Db 33 PACLY-----MGRQPPPPPPPPQFTSSLSGLSQSGFPDIPSYEVPPPLASDDDPAGAHLLHHL 87
Qy 78 AGFGGLSSQGVYVGPVGSFSKAGNEYPTRTR-----NCWADTGQDWRGSRAPC 126
Db 88 PAQLGLAHPPP--GFFPNGTEPGGLEPNNRVQLPFPWMKSTKAHAW--KGQ-WAGGA--- 139
Qy 127 GNTDPDPLSDTIHKKHTRPTFTGHQIFALEKTEQTKYLAGEPARARLAYSLGTMESQVKV 186
Db 140 -YTAEP-----EENKTRTAYTRAQLLEKEFLFNKYSRPRRVELAVMLNLTHERHIKI 193
Qy 187 WFQNRRTKWKKSALBESSSTPRAPGGA---SGDRAASENEDDEYNKPLDP 234
Db 194 WFQNRMKWKEEDKKSSTGTPSGGGGEEPEQDCAVTSGBELLAVPPLPP 244

RESULT 10
US-10-654-102-61
; Sequence 61, Application US/10654102
; Publication No. US20040132679A1
; GENERAL INFORMATION:
; APPLICANT: CHAN, LAWRENCE
; APPLICANT: KOJIMA, HIDETO
; TITLE OF INVENTION: INDUCTION OF PANCREATIC ISLET FORMATION
; FILE REFERENCE: P024090S1
; CURRENT APPLICATION NUMBER: US/10/654,102
; CURRENT FILING DATE: 2003-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-654-102-61

Query Match 15.5%; Score 217; DB 16; Length 284;
Best Local Similarity 30.3%; Pred. No. 1.9e-12;
Matches 70; Conservative 26; Mismatches 95; Indels 40; Gaps 9;

Qy 25 PMCOYSVQNSFYKLSPPGLGPQ-----LAAAGTGHGTTILSRPVATPNSLLSGYPHV 77
Db 33 PACLY-----MGRQPPPPPPPPQFTSSLSGLSQSGFPDIPSYEVPPPLASDDDPAGAHLLHHL 87
Qy 78 AGFGGLSSQGVYVGPVGSFSKAGNEYPTRTR-----NCWADTGQDWRGSRAPC 126
Db 88 PAQLGLAHPPP--GFFPNGTEPGGLEPNNRVQLPFPWMKSTKAHAW--KGQ-WAGGA--- 139
Qy 127 GNTDPDPLSDTIHKKHTRPTFTGHQIFALEKTEQTKYLAGEPARARLAYSLGTMESQVKV 186
Db 140 -YTAEP-----EENKTRTAYTRAQLLEKEFLFNKYSRPRRVELAVMLNLTHERHIKI 193
Qy 187 WFQNRRTKWKKSALBESSSTPRAPGGA---SGDRAASENEDDEYNKPLDP 234
Db 194 WFQNRMKWKEEDKKSSTGTPSGGGGEEPEQDCAVTSGBELLAVPPLPP 244

RESULT 11
US-10-654-102-63

```


Publication No. US20020197678A1	
GENERAL INFORMATION:	
APPLICANT: Jesell, Thomas M.	
APPLICANT: Tanabe, Yasuto	
APPLICANT: William, Christopher	
TITLE OF INVENTION: Gene Encoding MNR2 and Uses Thereof	
FILE REFERENCE: 57477/jpw/wl	
CURRENT APPLICATION NUMBER: US/10/095,932	
CURRENT FILING DATE: 2002-03-11	
PRIOR APPLICATION NUMBER: US/09/162,524	
PRIOR FILING DATE: 1998-09-29	
NUMBER OF SEQ ID NOS: 4	
SOFTWARE: PatentIn Ver. 2.0	
SEQ ID NO 3	
LENGTH: 349	
TYPE: PRT	
ORGANISM: chick embryo	
US-10-095-932-3	
Query Match 14.4%; Score 200.5; DB 13; Length 349;	
Best Local Similarity 25.4%; Pred. No. 1e-10;	
Matches 69; Conservative 29; Mismatches 73; Indels 101; Gaps 10;	
QY 39 SPPGLGPGQ-----LAAGTGHGTDILSRVATPNSLLSGYP-----HVAGFG 81	
DB 67 SPPRLPAHCALLPKAAFLGGGPGG-----GHPQHAGLGLHPAGPG 108	
QY 82 GLSSQGVYGVQVSFSGKAGNEYPTFNCWADTGQDWRGSRPCGNTDPL----- 133	
DB 109 G----PGLYGHFVYG--YPALGGQHPSLSYSQVQG-----AHPAHSADPKLSAGTFQ 158	
QY 134 -----SDTIHKKKHTRPTFTGHQIFALEKTFEQTQKYLGP 168	
DB 159 LDQWLRASTAGMILPKMDFDGSQSNLLGKCRPRPTFTSQQLLEHQLKLNKILSRP 218	
QY 169 ERARLAYSIGMTESQVKVWFQNRRTKWR--KKSALPSSSTPRAPGGASGD----- 217	
DB 219 KRFEVATSLMLTETQVKIWFQNRMRKWRQKKAQEAQAEANEKGGGGEDKSGPRELL 278	
QY 218 -----RAASENEDDEYNKPLDPDSDDEK 240	
DB 279 LFGPEKGGGRRLRELDPSE---PEDEEEEEEE 307	
RESULT 28	
US-10-654-102-58	
Sequence 58, Application US/10654102	
Publication No. US20040132679A1	
GENERAL INFORMATION:	
APPLICANT: CHAN, LAWRENCE	
APPLICANT: KOJIMA, HIDEYO	
TITLE OF INVENTION: INDUCTION OF PANCREATIC ISLET FORMATION	
FILE REFERENCE: P02409US1	
CURRENT APPLICATION NUMBER: US/10/654,102	
CURRENT FILING DATE: 2003-09-03	
NUMBER OF SEQ ID NOS: 194	
SOFTWARE: PatentIn Ver. 2.1	
SEQ ID NO 58	
LENGTH: 283	
TYPE: PRT	
ORGANISM: Rattus norvegicus	
US-10-654-102-58	
Query Match 14.0%; Score 195; DB 16; Length 283;	
Best Local Similarity 29.9%; Pred. No. 2.6e-10;	
Matches 66; Conservative 23; Mismatches 86; Indels 46; Gaps 11;	
QY 25 PMQCVSQVNSFYKLSPPGLGPQLAA-----GTPHGITDILSRPVA--TPNSSLLSGYPH 76	
DB 33 PACLY-----MGRQPPPPPPPPFAGSLGTLEQSPDDISPYEVPPLADDPAGAHLL--HHH 85	
QY 77 VAGFGGLSSQGVYGVQVGSFKAGNEYPTTR-----NCWADTGQDWRGSGARP 125	
DB 86 LPAQLGLAHPPP--GPPPNGTGTGGLSPSRVHLPPFMKSTKAHAWK---SOWAGGA-- 138	
QY 126 CGNTPDPLSDTIHKKKHTRPTFTGHQIFALEKTFEQTQKYLGPERRARLAYSIGMTESQVK 185	
DB 139 --YAAEP-----EENKTRTAYTRAQLLEKEFLFNKVISRPRRVELAVMLNLTERHIK 191	
QY 186 VWFQNRRTKWRKKSALPSSSTPRAPGGASGDRAASENEDD 226	
DB 192 IWFQNRMRKWKKEEDKKRSSGT--TSGGGGGE-----EPEQD 226	
RESULT 30	
US-10-097-340-71	
Sequence 71, Application US/10097340	
Publication No. US20030087250A1	
GENERAL INFORMATION:	
APPLICANT: John MONAHAN	
APPLICANT: Manjula GANNAVAPARAPU	
APPLICANT: Sebastian HOERSCH	
APPLICANT: Shubhangi KAMATKAR	
APPLICANT: Steve G. KOVATS	
APPLICANT: Rachel E. MEYERS	
APPLICANT: Michael MORRISEY	
APPLICANT: Peter OLANDT	
APPLICANT: Ami SEN	
APPLICANT: Peter VEIBY	
APPLICANT: Gordon B. MILLS	
APPLICANT: Robert C. BAST, Jr.	
APPLICANT: Karen LU	
APPLICANT: Rosemarie SCHMANDT	
APPLICANT: Xumei ZHAO	
APPLICANT: Karen GLATT	
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,	


```

; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 626
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-2995-027-626

```

	68,	Conservative	30,	Mismatch	76,	Index	36,	Gaps
Qy	20	SEMKAPNCQYSVONSFYKLSPPGLGQLAAGTPHGITDILSRPVATPNSSLSGYPH---	76	:	:	:	:	:
Dd	31	SQESPTLPEBSATDSDY-YSPTG-----GAPHGYCS-----PTSASYGKALNPYYQYQH	78	:	:	:	:	:
Qy	77	-VAGFGGLSQGVY-----YGQVGGSFSAKAGNEYPRTRNCWADTGQDWGRGSARPCGNTPD	131	:	:	:	:	:
Dd	79	GVNGSAGSYPAKAYADYSYASSYHQYGGAYNRVPSATN-----QPEKEVTE	124	:	:	:	:	:
Qy	132	P-----LSDTHKKKHRTPTTFGHQIFALEKTFTQTKLAGPERARLAYSLGWTESOVKW	187	:	:	:	:	:
Dd	125	PEVRMWNGPKVKRPRTIYSSFLQAALQRRFQTKYLALPERAEAAASLGLTQTQVKIW	184	:	:	:	:	:
Qy	188	FQNRTKWRK-----KSALPESSSTPRA-----PGGAS	215	:	:	:	:	:
Dd	185	FONKRKIIMKNEMPPHPSPSSDDPMACNSPOSVAWEPOGSS	230	:	:	:	:	:

```

RESULT 32
US-10-094-1844
; Sequence 7449, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOPYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/350,435

```



```
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1844
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-1844

Query Match      13.6%; Score 190.5; DB 15; Length 252;
Best Local Similarity 38.3%; Pred. No. 6.1e-10;
Matches 46; Conservative 19; Mismatches 44; Indels 11; Gaps 3;

QY 127 GNTDPLSDTIH-----KKKHTRTFTTGHQIFALEKTFEOTKYLAGPERARLAYSLGWT 181
Db 136 GNDTSPESFLLHNLARKPKRTAFSPQLRLLEHAFKKNHYVVGAEKQLAHSLSL 195
QY 182 SOVKVWFONRTKWRKKSALPSSSTPRAPGGA---SGDRAASENDEYKNKPLDPSDD 238
Db 196 TQVKVWFONRTKFKRQKLEEGSDSQKKKGTHHNRWRIATKQASP-----EIDVTSDD 252

RESULT 33
US-10-012-456A-38
; Sequence 38, Application US/10012456A
; Publication No. US20030087243A1
; GENERAL INFORMATION:
; APPLICANT: The Johns Hopkins University
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: IMPW/P23071PC
; CURRENT APPLICATION NUMBER: US/10/012,456A
; CURRENT FILING DATE: 2001-12-12
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: VARIANT
; LOCATION: (4)
; OTHER INFORMATION: any amino acid
US-10-012-456A-38

Query Match      13.5%; Score 188; DB 14; Length 311;
Best Local Similarity 26.1%; Pred. No. 1.4e-09;
Matches 69; Conservative 35; Mismatches 86; Indels 74; Gaps 11;

QY 8 TFLNNTQLAQFSEMKAPMCQYVQNSFYKLSPPGLGPOL-----AAGTP-----HG 54
Db 56 TFWIDBI-----LSKETCDY-----FKLSYVCPSLVVRPKPLHSCCTGSPSLRAYP 103
QY 55 ITDILSRPVATPNSLLSGYPHVAGFGLSSQGVY-----GPQVGSFSGAGNEYPTR 107
Db 104 LLSVITRQ-PTVISHLVATPGIA--QALSCHQVTEAVSAEAPGGEALASSESETEQPTP 160
QY 108 TRNCWADTQDWRGSRAPCGNTPDPLSDTIHKKHTRTFTTGHQIFALEKTFEOTKYLAG 167
Db 161 RQK-----KPRRSRTITFTELQMLGLEKFKQKYLST 192
QY 168 PERARLAYSLGWTESQVKVWFONRTKWRKKSALPSSSTPRAPGG-----ASGDR 219
Db 193 PDLDLAQSLGLTQLQVTKTYQNRRMKW-KKVLKGGQEAPTKPKRPNKNSIPTSEEIE 251
QY 220 ASE--NEDDEYNKPLDPSDDDEKI 241
Db 252 AEEKMNSAQGQQLPSPSQQEEL 275

RESULT 34
US-10-012-456A-54
; Sequence 54, Application US/10012456A
; Publication No. US20030087243A1
; GENERAL INFORMATION:
; APPLICANT: The Johns Hopkins University
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: IMPW/P23071PC
; CURRENT APPLICATION NUMBER: US/10/012,456A
; CURRENT FILING DATE: 2001-12-12
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 54
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-456A-54

Query Match      13.4%; Score 187.5; DB 14; Length 254;
Best Local Similarity 27.3%; Pred. No. 1.2e-09;
Matches 67; Conservative 31; Mismatches 80; Indels 67; Gaps 10;

QY 27 CQYSVQNSFYKLSPPGLGPOL-----AAGTP-----HGITDILSRPVATPNSLLSG 73
Db 11 CDY-----FKLSYVCPSLVVRPKPLHSCCTGSPSLRAYPLLSVITRQ-PTVISHLVPA 64
QY 74 YPHVAGFGLSSQGVY-----GPQVGSFSGAGNEYPTRTRNCWADTQDWRGSRAPC 126
Db 65 TPGIA--QALSCHQVTEAVSAEAPGGEALASSESETEQPTPRQK-----106
QY 127 GNTDPLSDTIHKKHTRTFTTGHQIFALEKTFEOTKYLAGPERARLAYSLGWTESQVKV 186
Db 107 -----KPRRSRTITFTELQMLGLEKFKQKYLSTPDLDLAQSLGLTQLQVKT 154
QY 187 WFNQNRRTKWRKKSALPSSSTPRAPGG-----ASGDRAASE--NEDDEYNKPLDPS 236
Db 155 WYQNRMKW-KKVLKGGQEAPTKPKRPNKNSIPTSEEIEAEKMSQAQGOQLPSPQ 213
QY 237 DDEKI 241
Db 214 QGEEL 218

RESULT 35
US-10-322-281-672
; Sequence 672, Application US/10322281
; Publication No. US2004012672A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 672
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-322-281-672

Query Match      13.3%; Score 186; DB 16; Length 312;
Best Local Similarity 35.0%; Pred. No. 2.2e-09;
Matches 62; Conservative 27; Mismatches 72; Indels 16; Gaps 9;

QY 68 SSSLGYPHVAGFGLS--SQGVYGPQVGSFSGAGNEYPTRT-RNCWADTQDWRGSRAP 125
Db 24 SSLASAY---ADFSQCSQASGFQYNPIRTTFTG-ATSGCPSLTGSCSLGTLRDHQS--P 77
QY 126 CGNTPDPLSDTIH-----KKKHTRTFTTGHQIFALEKTFEOTKYLAGPERARLAYSLGWT 180
Db 78 YAAVEYKLF-TDHGOEKQRRIITFTTSAQLKELERVFAETHYPTDIYTBRELALKIDIT 136
```

```
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1844
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-1844

Query Match      13.6%; Score 190.5; DB 15; Length 252;
Best Local Similarity 38.3%; Pred. No. 6.1e-10;
Matches 46; Conservative 19; Mismatches 44; Indels 11; Gaps 3;

QY 127 GNTDPLSDTIH-----KKKHTRTFTTGHQIFALEKTFEOTKYLAGPERARLAYSLGWT 181
Db 136 GNDTSPESFLLHNLARKPKRTAFSPQLRLLEHAFKKNHYVVGAEKQLAHSLSL 195
QY 182 SOVKVWFONRTKWRKKSALPSSSTPRAPGGA---SGDRAASENDEYKNKPLDPSDD 238
Db 196 TQVKVWFONRTKFKRQKLEEGSDSQKKKGTHHNRWRIATKQASP-----EIDVTSDD 252

RESULT 33
US-10-012-456A-38
; Sequence 38, Application US/10012456A
; Publication No. US20030087243A1
; GENERAL INFORMATION:
; APPLICANT: The Johns Hopkins University
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: IMPW/P23071PC
; CURRENT APPLICATION NUMBER: US/10/012,456A
; CURRENT FILING DATE: 2001-12-12
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: VARIANT
; LOCATION: (4)
; OTHER INFORMATION: any amino acid
US-10-012-456A-38

Query Match      13.5%; Score 188; DB 14; Length 311;
Best Local Similarity 26.1%; Pred. No. 1.4e-09;
Matches 69; Conservative 35; Mismatches 86; Indels 74; Gaps 11;

QY 8 TFLNNTQLAQFSEMKAPMCQYVQNSFYKLSPPGLGPOL-----AAGTP-----HG 54
Db 56 TFWIDBI-----LSKETCDY-----FKLSYVCPSLVVRPKPLHSCCTGSPSLRAYP 103
QY 55 ITDILSRPVATPNSLLSGYPHVAGFGLSSQGVY-----GPQVGSFSGAGNEYPTR 107
Db 104 LLSVITRQ-PTVISHLVATPGIA--QALSCHQVTEAVSAEAPGGEALASSESETEQPTP 160
QY 108 TRNCWADTQDWRGSRAPCGNTPDPLSDTIHKKHTRTFTTGHQIFALEKTFEOTKYLAG 167
Db 161 RQK-----KPRRSRTITFTELQMLGLEKFKQKYLST 192
QY 168 PERARLAYSLGWTESQVKVWFONRTKWRKKSALPSSSTPRAPGG-----ASGDR 219
Db 193 PDLDLAQSLGLTQLQVTKTYQNRRMKW-KKVLKGGQEAPTKPKRPNKNSIPTSEEIE 251
QY 220 ASE--NEDDEYNKPLDPSDDDEKI 241
Db 252 AEEKMNSAQGQQLPSPSQQEEL 275

RESULT 34
US-10-012-456A-54
; Sequence 54, Application US/10012456A
; Publication No. US20030087243A1
; GENERAL INFORMATION:
; APPLICANT: The Johns Hopkins University
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: IMPW/P23071PC
; CURRENT APPLICATION NUMBER: US/10/012,456A
; CURRENT FILING DATE: 2001-12-12
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 54
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-456A-54

Query Match      13.4%; Score 187.5; DB 14; Length 254;
Best Local Similarity 27.3%; Pred. No. 1.2e-09;
Matches 67; Conservative 31; Mismatches 80; Indels 67; Gaps 10;

QY 27 CQYSVQNSFYKLSPPGLGPOL-----AAGTP-----HGITDILSRPVATPNSLLSG 73
Db 11 CDY-----FKLSYVCPSLVVRPKPLHSCCTGSPSLRAYPLLSVITRQ-PTVISHLVPA 64
QY 74 YPHVAGFGLSSQGVY-----GPQVGSFSGAGNEYPTRTRNCWADTQDWRGSRAPC 126
Db 65 TPGIA--QALSCHQVTEAVSAEAPGGEALASSESETEQPTPRQK-----106
QY 127 GNTDPLSDTIHKKHTRTFTTGHQIFALEKTFEOTKYLAGPERARLAYSLGWTESQVKV 186
Db 107 -----KPRRSRTITFTELQMLGLEKFKQKYLSTPDLDLAQSLGLTQLQVKT 154
QY 187 WFNQNRRTKWRKKSALPSSSTPRAPGG-----ASGDRAASE--NEDDEYNKPLDPS 236
Db 155 WYQNRMKW-KKVLKGGQEAPTKPKRPNKNSIPTSEEIEAEKMSQAQGOQLPSPQ 213
QY 237 DDEKI 241
Db 214 QGEEL 218

RESULT 35
US-10-322-281-672
; Sequence 672, Application US/10322281
; Publication No. US2004012672A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 672
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-322-281-672

Query Match      13.3%; Score 186; DB 16; Length 312;
Best Local Similarity 35.0%; Pred. No. 2.2e-09;
Matches 62; Conservative 27; Mismatches 72; Indels 16; Gaps 9;

QY 68 SSSLGYPHVAGFGLS--SQGVYGPQVGSFSGAGNEYPTRT-RNCWADTQDWRGSRAP 125
Db 24 SSLASAY---ADFSQCSQASGFQYNPIRTTFTG-ATSGCPSLTGSCSLGTLRDHQS--P 77
QY 126 CGNTPDPLSDTIH-----KKKHTRTFTTGHQIFALEKTFEOTKYLAGPERARLAYSLGWT 180
Db 78 YAAVEYKLF-TDHGOEKQRRIITFTTSAQLKELERVFAETHYPTDIYTBRELALKIDIT 136
```


QY 181 ESQVQVWFQNRRTKWRKSALEPSSSTPRAPGASGDRAASENEDDEYN-KPLDPDS 236
 DB 137 EARVQVWFQNRRAKFRKQER-AAAAAANKGSSGKSDSRDDSEKAKSTDPDS 192

RESULT 36
 US-10-322-281-669
 ; Sequence 669, Application US/10322281
 ; Publication No. US20040126762A1
 ; GENERAL INFORMATION:
 ; APPLICANT: David W. Morris
 ; APPLICANT: Marc S. Malandro
 ; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
 ; FILE REFERENCE: 529452001000
 ; CURRENT APPLICATION NUMBER: US/10/322,281
 ; CURRENT FILING DATE: 2002-12-17
 ; NUMBER OF SEQ ID NOS: 866
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 669
 ; LENGTH: 340
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-322-281-669

Query Match 13.3%; Score 186; DB 16; Length 340;
 Best Local Similarity 35.0%; Pred. No. 2.5e-09;
 Matches 62; Conservative 27; Mismatches 72; Indels 16; Gaps 9;

QY 68 SLLSGYPHVAGFGGLS-SQVYYPQVGSFSGKAGNEYPTRT-RNCWADTGDWRGSRP 125
 DB 52 SSLASAY---ADFSSCSQASGFQYNPRTTFG-ATSGCPSLTPGCSGLTLRDHQS--P 105

QY 126 CNTPPDPLSDTH-----KKGTRPTTGHQIFALEKTFEQTLYAGPERARLAYSLGWT 180
 DB 106 YAAVPYKLF-TDHGQEKQRRTTFTSAQLKELERVFPAETHYDPDIYTREELALKIDL 164

QY 181 ESQVQVWFQNRRTKWRKSALEPSSSTPRAPGASGDRAASENEDDEYN-KPLDPDS 236
 DB 165 EARVQVWFQNRRAKFRKQER-AAAAAANKGSSGKSDSRDDSEKAKSTDPDS 220

RESULT 37
 US-10-755-889-552
 ; Sequence 552, Application US/10755889
 ; Publication No. US20040171823A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bristol-Myers Squibb Company
 ; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
 ; FILE REFERENCE: D0284 NP
 ; CURRENT APPLICATION NUMBER: US/10/755,889
 ; CURRENT FILING DATE: 2004-01-13
 ; PRIOR APPLICATION NUMBER: U.S. 60/440,068
 ; PRIOR FILING DATE: 2003-01-14
 ; PRIOR APPLICATION NUMBER: U.S. 60/469,757
 ; PRIOR FILING DATE: 2003-05-12
 ; NUMBER OF SEQ ID NOS: 823
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 552
 ; LENGTH: 149
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-755-889-552

Query Match 13.3%; Score 185; DB 16; Length 149;
 Best Local Similarity 41.5%; Pred. No. 1e-09;
 Matches 44; Conservative 13; Mismatches 45; Indels 4; Gaps 2;

QY 133 LSDTHKKKHRTPTTGHQIFALEKTFEQTLYAGPERARLAYSLGWTESQVQVWFQNR 192
 DB 9 LQRLHKRGQGVRFSDQTIIEKKKFTQKYLSPERKRLAKMLQLSERQVKTWFQNR 68

QY 193 TWKWKSALEPSSSTPRAPGASGDRAASENED--DEYNKPLDPDS 236

DB 69 AKWRRLKQENQSN--KKEELESLSDCDQDQLPSEQNKGSALDS 112

RESULT 38
 US-10-408-501-6
 ; Sequence 6, Application US/10408501
 ; Publication No. US20030224424A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Garcia-Barcelo, Maria Mercedes
 ; APPLICANT: Sham, Mai Har
 ; APPLICANT: Tam, Pau Kwong Hang
 ; APPLICANT: Lui, Vincent Chi Hang
 ; APPLICANT: Chen, Benedict Ling Sze
 ; TITLE OF INVENTION: PHOX2B POLYMORPHISMS AS HIRSCHSPRUNG'S
 ; FILE REFERENCE: 9661-031-999
 ; CURRENT APPLICATION NUMBER: US/10/408,501
 ; CURRENT FILING DATE: 2003-04-07
 ; PRIOR APPLICATION NUMBER: 60/370,968
 ; PRIOR FILING DATE: 2002-04-08
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 6
 ; LENGTH: 309
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-408-501-6

Query Match 13.3%; Score 185; DB 15; Length 309;
 Best Local Similarity 34.6%; Pred. No. 2.8e-09;
 Matches 62; Conservative 27; Mismatches 72; Indels 18; Gaps 9;

QY 68 SLLSGYPHVAGFGGLS-SQVYYPQVGSFSGKAGNEYPTRT-RNCWADTGDWRGSRP 125
 DB 24 SSLASAY---ADFSSCSQASGFQYNPRTTFG-ATSGCPSLTPGCSGLTLRDHQS--P 77

QY 126 CNTPPDPLSDTH-----KKGTRPTTGHQIFALEKTFEQTLYAGPERARLAYSLG 178
 DB 78 YAAVPYKLF-TDHGGLNEKQRRTTFTSAQLKELERVFPAETHYDPDIYTREELALKID 136

QY 179 MTESQVQVWFQNRRTKWRKSALEPSSSTPRAPGASGDRAASENEDDEYN-KPLDPDS 236
 DB 137 LTEARVQVWFQNRRAKFRKQER-AAAAAANKGSSGKSDSRDDSEKAKSTDPDS 194

RESULT 39
 US-10-408-501-2
 ; Sequence 2, Application US/10408501
 ; Publication No. US20030224424A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Garcia-Barcelo, Maria Mercedes
 ; APPLICANT: Sham, Mai Har
 ; APPLICANT: Tam, Pau Kwong Hang
 ; APPLICANT: Lui, Vincent Chi Hang
 ; APPLICANT: Chen, Benedict Ling Sze
 ; TITLE OF INVENTION: PHOX2B POLYMORPHISMS AS HIRSCHSPRUNG'S
 ; FILE REFERENCE: 9661-031-999
 ; CURRENT APPLICATION NUMBER: US/10/408,501
 ; CURRENT FILING DATE: 2003-04-07
 ; PRIOR APPLICATION NUMBER: 60/370,968
 ; PRIOR FILING DATE: 2002-04-08
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 314
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-408-501-2

Query Match 13.3%; Score 185; DB 15; Length 314;
 Best Local Similarity 34.6%; Pred. No. 2.8e-09;

		Matches	62;	Conservative	27;	Mismatches	72;	Indels	18;	Gaps	9;
Qy	68	SSLLSGYPHVAGFGGLS-SQGVVYGPQVGSFSKAGNEYPTRT-RNCWADTGQDWGRSARP	125								
Db	24	SSLASAY---ADFSSCSQASGFQYNPIRTTFG-ATSCPSLTPGSCSLGTLRDHQSS--P	77								
Qy	126	CGNTPDPLSDTIH-----KKKHTRPTTGHQIFALEKTFEQTKYLAGPEARLAYSLG	178								
Db	78	YAAVPYKLF-TDHGGLNEKKQRIRITFTSAQLKELERVFAETHYPDIYTREELALKID	136								
Qy	179	MTESQVKVWFONRRTKWRKKSALPSSSTPRAPGGASGDRAAASENEDDEYN-KPLDPDS	236								
Db	137	LTEARVQVWFQNRRAFRKQER-AAAAAAAAGSSGKSDSRDDESKEAKSTDPS	194								

RESULT 40
US-10-891-585-2
; Sequence 2, Application US/10891585
; Publication No. US20050042657A1
; GENERAL INFORMATION:
; APPLICANT: WESE-MAYER, Debra E.
; APPLICANT: BERRY-KRAVIS, Elizabeth M.
; APPLICANT: ZHOU, Lili
; TITLE OF INVENTION: METHODS AND PRIMERS FOR DIAGNOSING IDIOPATHIC CONGENITAL CENTRAL
; TITLE OF INVENTION: HYPOVENTILATOIN SYNDROME
; FILE REFERENCE: 045684-0101
; CURRENT APPLICATION NUMBER: US/10/891,585
; CURRENT FILING DATE: 2004-07-15
; PRIOR APPLICATION NUMBER: US 60/488,105
; PRIOR FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-891-585-2

Query Match	13.3%;	Score 185;	DB 17;	Length 314;
Best Local Similarity	34.6%;	Pred. No. 2.8e-09;		
Matches	62;	Conservative 27;	Mismatches 72;	Indels 18; Gaps 9;
Qy	68	SSLLSGYPHVAGFGGLS-SQGVVYGPQVGSFSKAGNEYPTRT-RNCWADTGQDWGRSARP	125	
Db	24	SSLASAY----ADFSSCSQASGFQYNPIRTTFG-ATSCPSLTPGSCSLGTLRDHQSS--P	77	
Qy	126	CGNTPDPLSDTIH-----KKKHTRPTTGHQIFALEKTFEQTKYLAGPEARLAYSLG	178	
Db	78	YAAVPYKLF-TDHGGLNEKKQRIRITFTSAQLKELERVFAETHYPDIYTREELALKID	136	
Qy	179	MTESQVKVWFONRRTKWRKKSALPSSSTPRAPGGASGDRAAASENEDDEYN-KPLDPDS	236	
Db	137	LTEARVQVWFQNRRAFRKQER-AAAAAAAAGSSGKSDSRDDESKEAKSTDPS	194	

Search completed: April 13, 2005, 17:22:41
Job time : 201.27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2005, 17:02:44 ; Search time 55.7635 Seconds
(without alignments)
452.066 Million cell updates/sec

Title: US-09-998-861-13
Perfect score: 1396
Sequence: 1 MESNLQGTFLNNTQLAQFS.....LLLRKHRAAFSVLSLGHVSU 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	583	41.8	277	2 S35304	homeotic protein G
2	553	39.6	364	2 I48188	gene NKX6.1 protein
3	228	16.3	277	2 S78063	homeobox protein P
4	221.5	15.9	284	2 B41224	homeotic protein P
5	219.5	15.7	327	2 A41168	homeotic protein D
6	217.5	15.6	271	2 S30230	homeotic protein H
7	217	15.5	284	2 S39581	IPFI protein - mou
8	215.5	15.4	332	2 I57032	gene Tlx-1 protein
9	214.5	15.4	838	2 I45557	eyeless, long form
10	212.5	15.2	328	2 G02469	homeotic protein D
11	212.5	15.2	330	2 A08855	homeotic protein H
12	212.5	15.2	543	2 B39369	homeotic protein B
13	210	15.0	270	2 JN0767	homeobox protein H
14	209.5	15.0	405	4 A61181	homeotic protein H
15	208	14.9	332	2 JH0465	homeotic protein T
16	205.5	14.7	373	2 A47234	homeobox protein H
17	204.5	14.6	217	1 WJHU2C	homeotic protein H
18	204.5	14.6	307	2 A45581	Distal-less homeob
19	204	14.6	604	2 A39369	homeotic protein B
20	203.5	14.6	606	2 S13367	Om(1D) protein - f
21	203	14.5	217	1 WJMSX2	homeotic protein H
22	202	14.5	307	2 S37252	homeotic protein H
23	201.5	14.4	229	2 A28329	homeotic protein H
24	201	14.4	232	2 T22698	hypothetical prote
25	200	14.3	283	2 I50112	homeotic protein P
26	200	14.3	458	2 I48690	homeotic protein N
27	199	14.3	298	2 T37251	homeobox protein c
28	198.5	14.2	264	1 WJHU3E	homeotic protein H
29	198.5	14.2	285	2 I51412	hypothetical trans

30	198	14.2	488	2 A55180	homeotic protein H
31	197.5	14.1	335	2 I57036	gene Dlx protein - H
32	196.5	14.1	476	2 A39621	homeotic protein H
33	196	14.0	232	2 A61045	homeotic protein T
34	196	14.0	305	2 S41860	gene Nkx-1.1 prote
35	195.5	14.0	258	2 JG6307	homeobox protein B
36	195.5	14.0	273	2 C88429	protein ceh-43 [im
37	195.5	14.0	659	2 A36664	S59/2 homeotic pro
38	195	14.0	220	2 S01063	Hox 2 protein type
39	195	14.0	241	2 S60249	homeotic protein f
40	195	14.0	283	2 S42634	homeoprotein, IDX-
41	193.5	13.9	264	1 S35219	homeotic protein H
42	193.5	13.9	270	2 I50111	Dlx2 homeodomain P
43	193.5	13.9	278	2 A56570	homeobox protein D
44	192.5	13.8	288	2 B40722	homeotic protein L
45	191.5	13.7	236	2 A46305	homeobox protein X

ALIGNMENTS

RESULT 1

S35304
homeotic protein Gtx - mouse
C:Species: Mus musculus (house mouse)
C>Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 16-Aug-2004
C:Accession: S35304
R:Komuro, I.; Schalling, M.; Jahn, L.; Bodmer, R.; Jenkins, N.A.; Copeland, N.G.; Izumo, EMBO J. 12, 1387-1401, 1993
A>Title: Gtx: a novel murine homeobox-containing gene, expressed specifically in glial, ible promoter.
A:Reference number: S35304; MUID:93223679; PMID:8096811
A:Accession: S35304
A:Molecule type: mRNA
A:Residues: 1-277 <KOW>
A:Cross-references: EMBL:L08074
C:Genetics:
A:Map position: 7
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:149-205/Domain: homeobox homology <HOX>

Query Match	41.8%	Score	583	DB	2	Length	277
Best Local Similarity	49.6%	Pred. No.	5.2e-41				
Matches	132	Conservative	29	Mismatches	81	Indels	24
Gaps	8						
QY	1	MESNLQGTFLNNTOLA--	OFSEMKAPMCQSVQN--	SFYKLSPPGLGPQLAAGTPHGI	55		
Db	1	MDANRPQAFVLSAPLAALHNVAEMKTSIFPVALQGPAGFKTPALGSLGAQLPLGTPHGI	60				
QY	56	TDILSRPVATPNSSLLSGYPHVAGFGGLSSQGVYVGPQVG--	SFSKAGNEYPTTRNCW	112			
Db	61	SDILGRPVGAAGGLGSLRLNGLA--	SSAGVYFGAAAVARGYPKPLAELGPRPIEW	118			
QY	113	ADT--GQDWR-----	GSARPCGNTDPDPLSDTHKKKHTPTFTGHQIFALEKTFEQTLYL	165			
Db	119	FGVQGSFWRPRLGSAQAG--	VLDKDKKKHSRPTFSGQIQIFALEKTFEQTLYL	173			
QY	166	ASPERARLAYSIGMTESQVKVWFQNRRTKWRKKSALPSSSTPRAPGASGGRASSEN--	223				
Db	174	ASPERARLAYSIGMTESQVKVWFQNRRTKWRKRAHAEMASAKKKQDSDAEKLVKVGSDAE	233				
QY	224	EDDEYNKFLDPDSDDEKIRLLLRKHR	249				
Db	234	DDDEYNRFLDPSDDDEKIRLLLRKHR	259				

RESULT 2

I48188
gene NKX6.1 protein - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Aug-2004
C:Accession: I48188

R;Rudnick, A.; Ling, T.Y.; Odagiri, H.; Rutter, W.J.; German, M.S.
Proc. Natl. Acad. Sci. U.S.A. 91, 12203-12207, 1994
A;Title: Pancreatic beta cells express a diverse set of homeobox genes.
A;Reference number: 148105; MUID:95083670; PMID:7991607
A;Accession: 148188
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-364 <RES>
A;Cross-references: UNIPROT:Q60554; EMBL:X81409; NID:9587466; PIDN:CAA57166.1; PID:g5974
C;Genetics:
A;Gene: NKX6.1
C;Superfamily: homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;237-293/Domain: homeobox homology <HOX>

Query Match 39.6%; Score 553; DB 2; Length 364;
Best Local Similarity 41.3%; Pred. No. 2.2e-38;
Matches 147; Conservative 26; Mismatches 81; Indels 102; Gaps 13;

QY 1 MESNLQGTFLNNTOLA---QFSEMKAPMC-----QYVQNSFYKLSPP-----41
DB 7 MEGPRQSAFLLSPPPLAALHSAEMKTELYPATYPPPLTPGPPSSSSSSSSSSSPPLGAH 66
QY 42 ----GLGP-----OLAGTGHGTDILSRP-----VATPNS-----69
DB 67 NPGGLKPPAAGLSLSPQQLSAATPHGINDILSRFMPVSGAALPSASPGSSSSS 126
QY 70 -----LLSGYPHVAGFGLS-----SQGVYVGPQ-----93
DB 127 SSSASNTSASAAAAAASAPAGLLAGLR---FSSLSPPPPPGLYFSPSAAA 183
QY 94 ----VGSFSKAGNEYPTRNCWADTQD---WRGSARPCGNTPD---LSDTIHKKHTRP 145
DB 184 VAAVGRYPKPLAELPGRAPIFWGVQSPWDRALAC--TPHQSGILLDKDGGKGRHP 241
QY 146 TFGHGFIALEKTFEOTKYLAGEPARLAYSLGMSQVKVWFQNRRTKWRKKSALPSS 205
DB 242 TFGQGIIFALEKTFEOTKYLAGEPARLAYSLGMSQVKVWFQNRRTKWRKKAEMAT 301
QY 206 STPRAPGASGDRAASEN--EDVYNKPLDPDSDDEKIRLLKRAAFSVLSGA 259
DB 302 AKKQDSETERLKGTSSENEEDDYNKPLDPNSDDEKITQLLKHGKSSGGLLHA 357

RESULT 3
S78063
homeobox protein Prh - chicken
C;Species: Gallus gallus (chicken)
C;Date: 19-Nov-1997 #sequence_revision 05-Dec-1997 #text_change 16-Aug-2004
C;Accession: S78063
R;Crompton, M.R.; Bartlett, T.J.; MacGregor, A.D.; Manfioletti, G.; Buratti, E.; Giancotti
Nucleic Acids Res. 20, 5661-5667, 1992
A;Title: Identification of a novel vertebrate homeobox gene expressed in haematopoietic
A;Reference number: S26799; MUID:93087175; PMID:1360645
A;Accession: S78063
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-277 <CR>
A;Cross-references: UNIPROT:Q05502; EMBL:X64711; NID:g297086; PIDN:CAA45966.1; PID:g2970
C;Superfamily: homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;145-201/Domain: homeobox homology <HOX>

Query Match 16.3%; Score 228; DB 2; Length 277;
Best Local Similarity 30.0%; Pred. No. 1.6e-11;
Matches 79; Conservative 25; Mismatches 87; Indels 72; Gaps 11;

QY 25 PMCQVQNSFYKLSPPGLGQPLAAGTGHGTDILSRVATPNS-----SLLSGY-----74
DB 25 PLLQPAHTPTFYIEDILGRGP-AAAPAPSLPAPPPPTLPSPNSSTFTSLVAPRTPTYPEP 83
QY 75 -----PHVAGFGLSSQGVYVGPQVGSFKAGNEY-----PTRTRNCWADTQD 118

84 TPIHPAFSHLAATYG---TGAYAGP-LYSPFPAVDYTHALIRQDPLGLKPLLSPPFIQ- 138
QY 119 WRGSARPCGNTPDPLSDTIHKKHTRPFTTGHQIFALEKTFEOTKYLAGEPARLAYSLG 178
DB 139 -----RP-----LHKRGQGVQVRSNEQITTELEKKFETQKYLSPFPERKRLAKLLQ 182
QY 179 MTESQVKVWFQNRRTKWRKKSALP-----SSSTPRAPGASGD-----217
DB 183 LSERQVKVWFQNRRAKWRRLKQENQATKKEAECTGHGDPSPSGSPAGGGEAPQD 242
QY 218 -RAASENEDVYNKPLDPDSDDE 239
DB 243 SPSAASQEDPE--SDVSDSDQEQ 263

RESULT 4
B41224
homeotic protein pMUR10F - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 09-Jul-2004
C;Accession: B41224
R;Kennedy, M.A.; Gonzalez-Sarmiento, R.; Kees, U.R.; Lampert, F.; Dear, N.; Boehm, T.; R.
Proc. Natl. Acad. Sci. U.S.A. 88, 8900-8904, 1991
A;Title: HOX11, a homeobox-containing T-cell oncogene on human chromosome 10q24.
A;Reference number: A41224; MUID:92020958; PMID:1681546
A;Accession: B41224
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-284 <KEN>
A;Cross-references: UNIPROT:Q61663; GB:M75953; NID:g193843; PIDN:AAA37805.1; PID:g193844
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;158-214/Domain: homeobox homology <HOX>

Query Match 15.9%; Score 221.5; DB 2; Length 284;
Best Local Similarity 21.8%; Pred. No. 5.6e-11;
Matches 68; Conservative 24; Mismatches 74; Indels 49; Gaps 8;

QY 39 SPPGLGQPLAAGT---PHGITDI-LSRPVATPNSLLSGYPHVAGFGLSSQGVYVGPQV 94
DB 61 APAGSLASLPRGSGVGGVIRVFAHRFLFVPPPS--GAAPAVPGPSGLGAG-----111
QY 95 GSFSKAGNEYPTRNCWADTQDQW-----RGSARPCGNTPDPLSDTIHK 139
DB 112 ---GLAGLTFF-----WMSGRFAKDRLTALSPSGTTRICHYPQNRTPP-----K 156
QY 140 KKHTRPFTTGHQIFALEKTFEOTKYLAGEPARLAYSLGMSQVKVWFQNRRTKWRKKS 199
DB 157 RKKPTSFSSQVLELERRFLRQKYLASAERAALAKALRMTDAQVKTWFQNRRTKWRQT 216
QY 200 ALEPSSSTPRAPGASGDRAASENEDVYNKPLDP 234
DB 217 AEEREASRHRA-----GRLLHLQDQALPRPLRP 245

RESULT 5
A44168
homeotic protein distal-less - fruit fly (Drosophila melanogaster)
N;Alternate names: homeotic protein Dll
C;Species: Drosophila melanogaster
C;Date: 27-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A44168; S03309
R;Vachon, G.; Cohen, B.; Pfeifle, C.; McGuffin, M.E.; Botas, J.; Cohen, S.M.
Cell 71, 437-450, 1992
A;Title: Homeotic genes of the Bithorax complex repress limb development in the abdomen
A;Reference number: A44168; MUID:93045644; PMID:1358457
A;Accession: A44168
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-327 <VAC>
A;Cross-references: UNIPROT:P20009; GB:S47947; NID:g259301; PIDN:AAB24059.1; PID:g25930
A;Note: sequence extracted from NCBI backbone (NCBI:P.117711)
R;Cohen, S.M.; Broenner, G.; Kuettner, F.; Juergens, G.; Jaekle, H.

Nature 338, 432-434, 1989
A:Title: Distal-less encodes a homeodomain protein required for limb development in Drosophila
A:Reference number: S03909; MUID:89181930; PMID:2564639
A:Accession: S03909
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 124-184 <COH>
C:Genetics:
A:Gene: dll
A:Cross-references: FlyBase:FBgn0000157
A:Map position: 2R 60B5,6
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:125-181/Domain: homeobox homology <HOX>

Query Match 15.7%; Score 219.5; DB 2; Length 327;
Best Local Similarity 34.6%; Pred. No. 9.7e-11;
Matches 55; Conservative 29; Mismatches 50; Indels 25; Gaps 5;

QY 78 AGFGGLSSQGVYGVQVGS-----FSKAGNEYPTRTNCWADTGDWRGSRAPCGNTP 130
DB 43 AGYGIRSTYQHFGQGGQDSFPSPRSALGYFPFPMHQNSY--SGYHLGSAVAPPCASPP 100
QY 131 D---PLSDTIH-----KKHTRPTTGHQIFALEKTFEOTKYLAGPERARLAYS 176
DB 101 KDFSISDKCEDSLRVNGKKGKMKRPRTIYSSLSQLQLNRRFQRTQYLAUPERAELAAS 160
QY 177 LGMTESQVQVWFQNRRTKWR--KSALEPSSSTPRAPGG 213
DB 161 LGLTQTVQVWFQNRRTKWR--KSALEPSSSTPRAPGG 199

RESULT 6
S30230
homeotic protein Hex - mouse
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Aug-2004
C:Accession: S30230
R:Bedford, F.K.; Ashworth, A.; Enver, T.; Wiedemann, L.M.
Nucleic Acids Res. 21, 1245-1249, 1993
A:Title: HEX: a novel homeobox gene expressed during haematopoiesis and conserved between Drosophila and vertebrates
A:Reference number: S30230; MUID:93219088; PMID:8096636
A:Accession: S30230
A:Molecule type: DNA
A:Residues: 1-271 <BED>
A:Cross-references: UNIPROT:P43120; EMBL:X74342; NID:G288500; PIDN:CAA79729.1; PID:G288500
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:139-195/Domain: homeobox homology <HOX>

Query Match 15.6%; Score 217.5; DB 2; Length 271;
Best Local Similarity 30.9%; Pred. No. 1.1e-10;
Matches 82; Conservative 19; Mismatches 65; Indels 99; Gaps 12;

QY 46 QLAAGTGHGTTDILSR-----SVATPN-----SSLLSGY-----PH 76
DB 26 QPAHPTFFYDILGGAAPPTTLPSPNSSFTLSVSSYRTPVYPTFVHPAFSHHPA 85
QY 77 VA-----GFGGLSSQGVYGVQVGSFKAGNEY-----PRTTRNCWADTGDWR 120
DB 86 AALAAAYGSPGFGG-----PLYPPRTVNDYTHALLRHDPGLKPLWSPLQ--- 132
QY 121 GSAPCGNTPDPLSTIHKKHTRPTTGHQIFALEKTFEOTKYLAGPERARLAYS LGMT 180
DB 133 ---RE-----LHKRGQVRFNSDQTVLEKKFETQKYLSPPERKRLAKWLQLS 178
QY 181 ESQVQVWFQNRRTKWR-----KXSALEPSSSTPRAPG-----CASDR--- 218
DB 179 ERQVKTWFQNRRAKRRLLKQENPQSNKDALD-SLDTSCQQLDLPSEQNKGASLDRSQC 237
QY 219 ----AAASENEDDEYNKPLDPDSDDE 239
DB 238 SPSPASQBDPDSSEISDSQEVDI 262

RESULT 7

S39581
IPF1 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Aug-2004
C:Accession: S39581
R:Ohlsson, H.; Karlsson, K.; Edlund, T.
EMBO J. 12, 4251-4259, 1993
A:Title: IPF1, a homeodomain-containing transactivator of the insulin gene.
A:Reference number: S39581; MUID:94038907; PMID:7901001
A:Accession: S39581
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-284 <OHL>
A:Cross-references: UNIPROT:P52946; EMBL:X74342; NID:G414678; PIDN:CAA52389.1; PID:G414678
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:148-204/Domain: homeobox homology <HOX>

Query Match 15.5%; Score 217; DB 2; Length 284;
Best Local Similarity 30.3%; Pred. No. 1.3e-10;
Matches 70; Conservative 26; Mismatches 95; Indels 40; Gaps 9;

QY 25 PNCQSVQNSFYKLSPPGLGPO-----LAAGTGHGTTDILSRPVATPNSLLSGYPRV 77
DB 33 FACLY-----MGROPPPPPPQFTSSLSGLSQGSPDISPYEVPPLASDDPAGAHLHHHL 87
QY 78 AGFGGLSSQGVYGVQVGSFKAGNEYPTRTT-----NCWADTGDWRGSRAPC 126
DB 88 PAQLGLAHPPP--GFPNGTEFGGLEENRVQLPFPWMKSTKAAHAW--KGQ-WAGGA--- 139
QY 127 GNTDPPLSDTIHKKHTRPTTGHQIFALEKTFEOTKYLAGPERARLAYS LGMTESQV 186
DB 140 -YTAEP-----EENKTRTAVTRAQLLEKEFLPNKYISRRRVELAVMLNLTERRHIKI 193
QY 187 WFOQRRTKWRKKSALPSSSTPRAPGGA---SGDRAASENEDDEYNKPLDP 234
DB 194 WFOQRRTKWRKKSALPSSSTPRAPGGA---SGDRAASENEDDEYNKPLDP 244

RESULT 8
S157032
gene Tlx-1 protein - mouse
C:Species: Mus sp. (mouse)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 16-Aug-2004
C:Accession: I57032
R:Raju, K.; Tang, S.; Dube, I.D.; Kamel-Reid, S.; Bryce, D.M.; Breitman, M.L.
Mech. Dev. 44, 51-64, 1993
A:Title: Characterization and developmental expression of Tlx-1, the murine homolog of Drosophila Tbx-1
A:Reference number: I57032; MUID:94206842; PMID:7908826
A:Accession: I57032
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-332 <RES>
A:Cross-references: GB:S70632; NID:G546379; PIDN:AAB30542.1; PID:G546380
C:Genetics:
A:Gene: Tlx-1
A:Introns: 192/1; 259/2
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:204-260/Domain: homeobox homology <HOX>

Query Match 15.4%; Score 215.5; DB 2; Length 332;
Best Local Similarity 30.0%; Pred. No. 2.1e-10;
Matches 73; Conservative 30; Mismatches 77; Indels 63; Gaps 9;

QY 36 YKLSPPGLGQAAAGTGHGTTDILSRPVATPNSLLSGYPHVAGFGG-----LSRSG 87
DB 74 YGAGGPG--GPGGPGAGGGGACSMGLPGSYNNMDLAGPGGPGGGGGGAARRALSAAG 132
QY 88 VY-----YGPQVGSFSK-----AGNEYPTRTNCWADTGDWR- 119


```
Db 133 VIRVAHRLAGVAHPOPLATGLPTVPSVPAVGVNVLGLTTP-----WVESNRYT 186
Qy 120 --RGSARPCGNTDPLSDTIHKKHHTPTFTGHQIFALEKTFEQTKYLAGPERARLAYS 177
Db 187 KDRFTGLPYQNTTP-----KKKKPRTSFTRLQICELEKRFHQKYLASAERAAALAKAL 240
Qy 178 GNTESQVQVWFQNRRTKWKKSALBPSSTPRAGSGADRAASENEDDEYNK-----P 231
Db 241 KMTDAQVKTWFQNRRTKWRRTAEERAEBSQA-----NRILLOQOEAFQKSLAOLPLP 294
Qy 232 LDP 234
Db 295 ADP 297

RESULT 9
145557
eyeless, long form - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 09-Mar-1996 #sequence_revision 09-Mar-1996 #text_change 16-Aug-2004
C:Accession: I45557
R:Quiring, R.; Wallidorf, U.; Kloter, U.; Gehring, W.J.
Science 265, 785-789, 1994
A:Title: Homology of the eyeless gene of Drosophila to the Small eye gene in mice and A
A:Reference number: A54584; MUID:942323757; PMID:7914031
A:Accession: I45557
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-838 <RES>
A:Cross-references: EMBL:X79493; NID:9641809; PIDN:CAA56038.1; PID:9641810
C:Genetics:
A:Gene: FlyBase:ey
A:Cross-references: FlyBase:FBgn000558
A:Introns: 37/1; 92/2; 152/3; 371/1; 429/1; 521/1; 639/2
C:Superfamily: homeobox homolog; paired box homology
C:Keywords: alternative splicing; DNA binding; homeobox; nucleus; transcription regulati
F:37-161/Domain: paired box homology <PBH>
F:412-468/Domain: homeobox homology <HOX>

Query Match 15.4%; Score 214.5; DB 2; Length 838;
Best Local Similarity 27.0%; Pred. No. 8.1e-10;
Matches 64; Conservative 32; Mismatches 82; Indels 59; Gaps 6;

Qy 12 NNTQLAQFSEMKAPMCQYQVNSFYKLS-----PGLGPOLAAAGTPHGITDILSRPVATPN 67
Db 288 NHOALQHQOQSWPPRPHYS--GSWPTSLSEIPISAPNIAVSATAYAGPSLAHSLSPN 345
Qy 68 SLLSGYPHVAGFGLSSQGVYGVQVGSFKAGNEYPRTRNCWADT----- 115
Db 346 -----DIKSLASTGHQ-----RNCVPVATEDIHLKKELDGH 375
Qy 116 -----QDWRGSRPCGNTDPLSDTIHKK--HTRPFTTGHQIFALEKTFEQTKY 164
Db 376 QSDETSGGSGNSGANGNIGTDDQALILKRLQNRNFTNDQIDSLKEPERTHY 435
Qy 165 LAGPERARLAYSIGMTESQVQVWFQNRRTKWKKSALBPSSTPRAPGSGADRAAS 221
Db 436 PDVFARERLAGLGLPEARIQVWFNSNRKRWREKLNQRRTPNSTGASATSSSTS 492

RESULT 10
G02469
homeotic protein DLX-2 - human
C:Species: Homo sapiens (man)
C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 16-Aug-2004
C:Accession: G02469; FN0670
R:Rubenstein, J.L.R.
submitted to the EMBL Data Library, March 1996
A:Reference number: H01328
A:Accession: G02469
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: mRNA

A;Residues: 1-328 <RUB>
A;Cross-references: UNIPROT:Q07687; EMBL:U51003; NID:gl477591; PIDN:AAB40902.1; PID:9147.
R;Seleki, D.J.; Thomas, N.E.; Coleman, P.D.; Rogers, K.E.
Gene 132, 301-303, 1993
A;Title: The human brain homeogene, DLX-2: cDNA sequence and alignment with the murine h
A;Reference number: FN0670; MUID:94040778; PMID:7901126
A:Accession: FN0670
A:Molecule type: mRNA
A;Residues: 85-328 <SEL>
A;Cross-references: GB:L07919; NID:9306709; PIDN:AAAL9663.1; PID:9306710
C:Genetics:
A;Gene: Dlx2
A;Introns: 134/1; 195/3
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:153-209/Domain: homeobox homology <HOX>

Query Match 15.2%; Score 212.5; DB 2; Length 328;
Best Local Similarity 28.9%; Pred. No. 3.7e-10;
Matches 68; Conservative 42; Mismatches 86; Indels 39; Gaps 10;

Qy 19 FSEMKAPM--COYSVONSFYKLSPPGLGPOLAAAGTPHGITDILSRPVATP-----NS 68
Db 5 FDSLIVADMHSTQIAASTYHQHQOPSGGGAGFGGNSSSSSSLHKQESPTLPVSTATDS 64
Qy 69 SLLSGYPHVAGFGLSSQGVYGVQVGSF---SKAGNEYPRTRNCWADTG-----Q 117
Db 65 SYVTNQHPAGGGG--GGGSPYA--HMGSYQYQASGLNNVFSYSAKSY-DLGYTAAYTSYA 120
Qy 118 DWRGSRPCGNTDPP-----LSDTIHKKHHTPTFTGHQIFALEKTFEQTKYLAGP 168
Db 121 PYGTSSSPANPEKEDLEPEIRIVNGKPKVKPRPTIYSSFQLAALORFQKTYLALP 180
Qy 169 ERARLAYSIGMTESQVQVWFQNRRTKWK--KSALEPSSSTPRAPGSGADRAAS 221
Db 181 ERAELASLGLTQVQKWFQNRKSKPKWKSGEIPSE---QHPGASASPPCAS 232

RESULT 11
A40855
homeotic protein Hox 11 - human
N:Alternate names: tcl-3 proto-oncogene
C:Species: Homo sapiens (man)
C>Date: 28-Feb-1992 #sequence_revision 28-Feb-1992 #text_change 09-Jul-2004
R:Atatano, M.; Roberts, C.W.M.; Minden, M.; Crist, W.M.; Korsmeyer, S.J.
Science 253, 79-82, 1991
A:Title: Deregulation of a homeobox gene, HOX11, by the t(10;14) in T cell leukemia.
A:Reference number: A40855; MUID:91289163; PMID:1676542
A:Accession: A40855
A:Molecule type: mRNA
A;Residues: 1-330 <HAT>
A;Cross-references: UNIPROT:P31314; GB:S38742; NID:9232582; PIDN:AAB19293.1; PID:923258
R;Lu, M.; Gong, Z.; Shen, W.; Ho, A.D.
EMBO J. 10, 2905-2910, 1991
A:Title: The tcl-3 proto-oncogene altered by chromosomal translocation in T-cell leukemia
A:Reference number: S17311; MUID:92007734; PMID:1717256
A:Accession: S17311
A:Molecule type: mRNA
A;Residues: 1-275,'L',277-330 <LUM>
A;Cross-references: EMBL:M62626; NID:9339199; PIDN:AAA36719.1; PID:9387683
R;Kennedy, M.A.; Gonzalez-Sarmiento, R.; Kees, U.R.; Lampert, F.; Dear, N.; Boehm, T.;
Proc. Natl. Acad. Sci. U.S.A. 88, 8900-8904, 1991
A:Title: HOX11, a homeobox-containing T-cell oncogene on human chromosome 10q24.
A:Reference number: A41224; MUID:92020958; PMID:1681546
A:Accession: A41224
A:Molecule type: DNA
A;Residues: 1-189,'VALSPFTVTRR',190-275,'L',277-330 <KEN>
A;Cross-references: GB:M75952
A>Note: the authors suggest a long form with twelve additional amino acids from possible
C:Genetics:
A:Gene: GDB:HOX11
A;Cross-references: GDB:119607; OMIM:186770
```


A:Map position: 10q24-10q24
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; leukemia; nucleus; proto-oncogene; transcription regulation
F:202-258/Domain: homeobox homology <HOX>

Query Match 15.2%; Score 212.5; DB 2; Length 330;
Best Local Similarity 30.2%; Pred. No. 3.7e-10;
Matches 74; Conservative 29; Mismatches 73; Indels 69; Gaps 11;
QY 36 YKLSPPLGQPOLAAGTHGHTDILSRVATPNSL---LSGYPHVAGFGG-----LSS 85
DB 74 YGTGFGF-GP-----GPGAGGCGACSMGPLTGSYNNMALAGGPGGCGGGGGGAGALSA 128
QY 86 QVY-----YGPVGSFSK-----AGNYPTTRNCWADTQOD 118
DB 129 AGVIRPAPRPLAGAVAHQPLATGLTPVSPAMPGVNNLTGLTFP-----WMESNRR 182
QY 119 W---RGSARPCGNTDPLSDTHKKHTRPTTGTGHOIFALEKTFPTQTKYLACPERARLAY 175
DB 183 YTKDFTGHPYQNRTPP-----KKKPKTSFTRIQICELEKRFHQRKYLASAERAAALAK 236
QY 176 SIGMTESQVWVFQNRRTKWRKKSALPSSSTPRAPGGASGDRAASENDDYNK----- 230
DB 237 ALKMTDAQVKTWFQNRRTKWRQTAEREAEERQQA-----NRILRQLQQAFAQKSLAQ 290
QY 231 -PLDP 234
DB 291 LPADP 295

RESULT 12

B39369
homeotic protein BarH1 - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C>Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 17-Oct-1997
C:Accession: B39369
R:Kojima, T.; Ishimaru, S.; Higashijima, S.; Takayama, E.; Akimaru, H.; Sone, M.; Emori, Proc. Natl. Acad. Sci. U.S.A. 88, 4343-4347, 1991
A:Title: Identification of a different-type homeobox gene, BarH1, possibly causing Bar (C) syndrome in humans
A:Reference number: A39369; MUID:91239564; PMID:1674606
A:Accession: B39369
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-543 <KO>
A:Cross-references: GB:M59965; GB:M73259; GB:M73078; GB:M73079
C:Genetics:
A:Gene: FlyBase:B-H1
A:Cross-references: FlyBase:FBgn0011758
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:300-356/Domain: homeobox homology <HOX>

Query Match 15.2%; Score 212.5; DB 2; Length 543;
Best Local Similarity 29.0%; Pred. No. 6.9e-10;
Matches 54; Conservative 33; Mismatches 74; Indels 25; Gaps 4;
QY 40 PPGLG-----PQLAAGTHGHTDILSRVATPNSLSGYPHVAGFGGLSSQ----- 86
DB 176 PPTAGGGLNVAQYAAQAQHHYAAAAAARNNAAAAAARAAAAAGVAAAPPVGGVD 235
QY 87 -GVYGPQVGSFSKAGNEYPFTRNC-----WADTQDWRGSRAPCGNTDPLSD 135
DB 236 GGVGLAPPAGGLDSDSDSYHEENEDCDGNMDDHSCVNGGKDDGNSVKSSTSD-MSG 294
QY 136 THKKHTRPTTGTGHOIFALEKTFPTQTKYLACPERARLAYSLGMTESQVWVFQNRRTKW 195
DB 295 LSKKORKARTPTDQLQLEKSFQRKYLVSQVQELAHKLDLSDCCQVKTWQNRRTKW 354
QY 196 RKKSAL 201
DB 355 KQQTAV 360

RESULT 13

JN0767
homeobox protein HEX - human
C:Species: *Homo sapiens* (man)
C>Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 16-Aug-2004
C:Accession: JN0767; S26799; S78048; S30231
R:Hromas, R.; Radich, J.; Collins, S. 1993
Biochem. Biophys. Res. Commun. 195, 976-983, 1993
A:Title: PCR cloning of an orphan homeobox gene (PRH) preferentially expressed in myeloid leukemia cell lines
A:Reference number: JN0767; MUID:93384629; PMID:8103988
A:Accession: JN0767
A:Molecule type: DNA
A:Residues: 1-270 <AC>
A:Cross-references: UNIPROT:Q03014; GB:L16499; NID:9292404; PIDN:AAA02988.1; PID:G9292405
R:Crompton, M.R.; Bartlett, T.J.; MacGregor, A.D.; Manfioletti, G.; Buratti, E.; Giancoli, A. 1992
Nucleic Acids Res. 20, 5661-5667, 1992
A:Title: Identification of a novel vertebrate homeobox gene expressed in haematopoietic cells
A:Reference number: S26799; MUID:93087175; PMID:1360645
A:Accession: S26799
A:Molecule type: mRNA
A:Residues: 1-114, 'L', 116-270 <CRO>
A:Cross-references: EMBL:X67235; NID:932547; PIDN:CAA47661.1; PID:G32548
R:Bedford, F.K.
submitted to the EMBL Data Library, February 1993
A:Reference number: S78048
A:Accession: S78048
A:Molecule type: DNA
A:Residues: 122-270 <BED>
A:Cross-references: EMBL:Z21533; NID:932068; PIDN:CAA79730.1; PID:G32069
R:Bedford, F.K.; Ashworth, A.; Enver, T.; Wiedemann, L.M.
Nucleic Acids Res. 21, 1245-1249, 1993
A:Title: HEX: a novel homeobox gene expressed during haematopoiesis and conserved between vertebrates
A:Reference number: S30230; MUID:93219088; PMID:8096636
A:Accession: S30231
A:Molecule type: DNA
A:Residues: 136-195 <BEW>
A:Cross-references: EMBL:Z21533
C:Comment: This protein is an important regulator of normal hematopoiesis.
C:Genetics:
A:Gene: HEX
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:138-194/Domain: homeobox homology <HOX>

Query Match 15.0%; Score 210; DB 2; Length 270;
Best Local Similarity 36.1%; Pred. No. 4.7e-10;
Matches 78; Conservative 20; Mismatches 86; Indels 32; Gaps 11;
QY 46 QLAAGTHGHTDILSR-PVA-----TPNS---SLSGY-----PHYAGGGLSS 85
DB 25 QPAHPTPFYIEDILGRGAAPTAPTLPSNPSSTLSVSPYRTVYETPIHPAFSHHSA 94
QY 86 QGV--YVGPQVGSFSKAGNEYP-TRTRNCWADTQDWRGSRAPCGNTDPLSDTHKKKH 142
DB 85 AALAAAYGP--GGP--GGPLIPFPTVNDYTHALVRHDPGLKPLWSFP-LQRPILKRG 139
QY 143 TRPTTGTGHOIFALEKTFPTQTKYLACPERARLAYSLGMTESQVWVFQNRRTKWRKKSAL 202
DB 140 GQVRFNSDQITTELEKFKFQKYLSPERKRAKMLQLSERQVKTWFQNRRAKWRRLKQEN 199
QY 203 PSSSTPRAPGASGDRAASENED--DEYNKPLDPS 236
DB 200 PQSN--KKBELESLSDDSCDQDLFPSEQNKASLDS 233

RESULT 14

A61181
homeotic protein HOX11-derived mutant fusion protein - human
C:Species: *Homo sapiens* (man)
C>Date: 10-Mar-1994 #sequence_revision 09-May-1996 #text_change 13-Aug-1999
C:Accession: A61181
R:Dube, I.D.; Kamel-Reid, S.; Yuan, C.C.; Lu, M.; Wu, X.; Corpus, G.; Raimondi, S.C.; Blood 78, 2996-3003, 1991

A>Title: A novel human homeobox gene lies at the chromosome 10 breakpoint in lymphoid ne
A:Reference number: A61181; MUID:92063028; PMID:1683261
A:Accession: A61181
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-405 <DUB>
A:CROSS-references: GB:X16734; NID:g525225
C:Genetics:
A:Gene: HOX11/unknown
A:Map position: 10q24/unknown
C:Keywords: homeobox; fusion protein

Query Match 15.0%; Score 209.5; DB 4; Length 405;
Best Local Similarity 32.1%; Pred. No. 8.6e-10;
Matches 72; Conservative 22; Mismatches 65; Indels 55; Gaps 10;
QY 36 YKLSPPLGQLAAGTGHITDIILSRPVATPNSL---LSGYPHVA-----GFGGLSSQ 86
DB 74 YGTGGFG-GP---GGPAGGAGCAGMPLTGSYNVNMALAGGPGLAAGGAGAGAGAGAGAG 128
QY 87 GYV-----YGPQVGSFK-----AGNEYPTETRCNWDATGQDW 119
DB 129 GVIRVFAHPLAGAVAHFQPLATGLTPVSPMPGVNLTGLTFP-----WMESNRY 182
QY 120 ---RGSARPCGNTDPDLSDTIHKKHTRPTFTGHQIFALEKTFEQTXYLAGPERALAYS 176
DB 183 TKDRFTGHYQNRTP-----KKKKPRTSFTLQICELEKRFHQKYLASAERAAALAKA 236
QY 177 LGMTESQVQKWFQNRRTKWRK-----KSALEPSSSTPRAP 211
DB 237 LKMTAQVKTWFQNRRTKWRRLRRNGRDRQATRSSSSCSRRP 280

RESULT 15
JH0465
homeotic protein Tes-1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 31-Mar-1992 #sequence_revision 15-Jan-1993 #text_change 16-Aug-2004
C:Accession: JH0465; A38859
R:Porteus, M.H.; Bulfone, A.; Ciaranello, R.D.; Rubenstein, J.L.R.
Neuron 7, 221-229, 1991
A>Title: Isolation and characterization of a novel cDNA clone encoding a homeodomain tha
A:Reference number: JH0465; MUID:91337450; PMID:1678612
A:Accession: JH0465
A:Molecule type: mRNA
A:Residues: 1-270, RAAPR, 277-278, 'WETTRG', 298, 'TR', 301, 'RA', 304, 'LH', 307, 'CRPOR', 313, '
A:CROSS-references: UNIPROT:P40764
A:Experimental source: brain
A>Note: this sequence has been revised in reference A38859
R:Porteus, M.H.; Bulfone, A.; Ciaranello, R.D.; Rubenstein, J.L.R.
Neuron 9, 186a, 1992
A:Reference number: A38859
A:Contents: erratum
A:Accession: A38859
A:Molecule type: mRNA
A:Residues: 261-332 <POR>
A:CROSS-references: GB:M80540
C:Comment: This protein is the mouse homolog of distal-less, a gene required for limb de
C:Genetics:
A:Gene: Tes-1
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:156-212/Domain: homeobox homology <HOX>

Query Match 14.9%; Score 208; DB 2; Length 332;
Best Local Similarity 30.5%; Pred. No. 8.9e-10;
Matches 73; Conservative 40; Mismatches 82; Indels 44; Gaps 12;
QY 19 FSEKAPM---CQYVQNSQVFKLSPP-----GLGP-----QLAAGTGHITDIILSRPVAT- 65
DB 5 FDSLVADMHSTQITASTYHQHQPPSGAGAGPGGNSSSSSSSSLHKKPQESPTLPVSTA 64
QY 66 PNSSLLSGYPHVAGFG-----LSSQGVYGPQV-----SFSKAGNEYP-----TRTNCWA 113

Db 65 TDSYYVNTQOHPPAGGCGGASPYAHMGSYQYHAGSLNNVSY- AKSSYDLGYTAAYTSA 123
QY 114 DTQDWRGSGARPCGNTDPD-----LSDTIHKKHTRPTFTGHQIFALEKTFEQTXY 164
DB 124 PYGT-----SSSPVNNFEDKEDLEPEIRIVNGPKPKVKRPTIYSSQLAALQRRFOKTOY 179
QY 165 LAGPERALAYS LGMTESQVQKWFQNRRTKWRK--KSALEPSSSTPRAPGASGDRAAS 221
DB 180 LALPERAELAAASLGITQTQVKIWFQNRSKFKMKWSGEIP---TQHPGASASPPCAS 235

RESULT 16
A47234
homeobox protein H6 - human
C:Species: Homo sapiens (man)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A47234
R:Stadler, H.S.; Padanilam, B.J.; Buetow, K.; Murray, J.C.; Solursh, M.
Proc. Natl. Acad. Sci. U.S.A. 89, 11579-11583, 1992
A>Title: Identification and genetic mapping of a homeobox gene to the 4p16.1 region of h
A:Reference number: A47234; MUID:93087572; PMID:1360670
A:Accession: A47234
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-373 <STA>
A:CROSS-references: UNIPROT:Q9NP08
A:Experimental source: embryo craniofacial region
A>Note: sequence extracted from NCBI backbone (NCBIN:119953, NCBI:119955)
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:201-257/Domain: homeobox homology <HOX>

Query Match 14.7%; Score 205.5; DB 2; Length 373;
Best Local Similarity 32.1%; Pred. No. 1.7e-09;
Matches 63; Conservative 18; Mismatches 72; Indels 43; Gaps 6;
QY 40 PPGIGPOLAAGTGHITDIILSRPVATPNSLSSGYPHVAG--FGGLS-----SOGVYGP 92
DB 97 PPGGPPFPALGC-----GGAGRWPYRAHGGYGGGLSPDTSDRDSPETGE 140
QY 93 QVGSFSKAGNEYPTETRCNWDATGQDWGRGSRPCG-----NTPDPLSDTI----- 137
DB 141 EMGRAEGANPRGPGRS-----GAAGSGAGGANPGRHGGSGELAEVPAAGETGGVGVG 196
QY 138 -HKKHTRPTFTGHQIFALEKTFEQTXYLAGPERALAYS LGMTESQVQKWFQNRRTKWR 196
DB 197 GGRKKKTITVFSRSQVFOLESTFDLKYLSAERAGLAASLQLTETQVKIWFQNRNRKWK 256
QY 197 KSALEPSSSTPRAPG 212
DB 257 RHVAAELEAAASLSPPG 272

RESULT 17
WHU2C
homeotic protein Hox B7 - human
N:Alternate names: homeotic protein ci; homeotic protein Hox 2C; TATAA binding protein
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C:Accession: A28030; S15535; A44934
R:Simeone, A.; Mavilio, F.; Acampora, D.; Giampaolo, A.; Faiella, A.; Zappavigna, V.; D.
Proc. Natl. Acad. Sci. U.S.A. 84, 4914-4918, 1987
A>Title: Two human homeobox genes, ci and c8: structure analysis and expression in embr
A:Reference number: A28030; MUID:87260899; PMID:2885844
A:Accession: A28030
A:Molecule type: mRNA
A:Residues: 1-217 <SIM>
A:CROSS-references: UNIPROT:P09629; GB:M16937
A>Note: the authors translated the codon GGC for residue 53 as Ala
R:Boncinelli, E.; Acampora, D.; Pannese, M.; d'Esposito, M.; Somma, R.; Gaudino, G.; St
Genome 31, 745-756, 1989
A>Title: Organization of human class I homeobox genes.

Matches 49; Conservative 25; Mismatches 48; Indels 25; Gaps 3;

QY 78 ACFGGLSSGVVYGVQVFSKAGNEYFTRTENCWADT----- 115
DB 248 AGCGGGGGUGV-GGAPAGALDDSDYIHEENEDCDSDGGGAGGGGSGHMDHDSVCSN 306
QY 116 -GDWRGSRPCCGNTPDPLSDTIHKKKHTPTFTTGHQI'FALEKTPQYKYLAGEPARLA 174
DB 307 GCKDDGNSIKSGSTSD-MSGLSKQKQKARTAFTHQTLQLEKSPERQKYLVSQVROELA 365
QY 175 YSLGHTESQVKVFWQNRRTKWKKSAL 201
DB 366 HKLILSDCQKTYQNRRTKWKROTAV 392

RESULT 21

WJMSX2

homeotic protein Hox B7 - mouse

N;Alternate names: homeotic protein Hox 2.3

C;Species: Mus musculus (house mouse)

C;Date: 30-Jun-1991 #sequence revision 30-Jun-1991 #text change 09-Jul-2004

C;Accession: A26846; E27176; A29585; S00988; I48411; S01887

R;Weilink, F.; de Laaf, R.; Verrilizer, P.; Deestree, O.; Kroezen, V.; Hilken, J.; Desch

Nucleic Acids Res. 15, 6773-6786, 1987

A;Title: A mouse homeobox containing gene on chromosome 11: sequence and tissue-specific

A;Reference number: A26846; MUID:88015526; PMID:2889183

A;Accession: A26846

A;Molecule type: DNA

A;Residues: 1-217 <ME1>

A;Cross-references: UNIPROT:P09024; GB:Y00436; NID:G51387; PIDN:CAA68494.1; PID:G861031

A;Accession: E26846

A;Molecule type: mRNA

A;Residues: 1-217 <ME2>

A;Cross-references: EMBL:Y00436; NID:G51387; PIDN:CAA68494.1; PID:G861031

R;Hart, C.P.; Fainsod, A.; Ruddle, F.H.

Genomics 1, 182-195, 1987

A;Title: Sequence analysis of the murine Hox-2.2, -2.3, and -2.4 homeo boxes: evolutionary

A;Reference number: A27176; MUID:88085193; PMID:2891608

A;Accession: E27176

A;Molecule type: DNA

A;Residues: 134-210, 'R', 212-217 <HAR>

A;Cross-references: EMBL:M18400

R;Lonai, P.; Arman, E.; Czosnek, H.; Ruddle, F.H.; Blatt, C.

DNA 6, 409-418, 1987

A;Title: New murine homeoboxes: structure, chromosomal assignment, and differential exp

A;Reference number: A29585; MUID:88054465; PMID:2890503

A;Accession: A29585

A;Molecule type: DNA

A;Residues: 'LCV', 134-185, 'G', 187-205, 'H', 207-210, 'A', 212-217 <ION>

A;Cross-references: EMBL:M18167

A;Note: the authors translated the codon CAG for residue 186 as Gly

R;Kongswan, K.; Webb, E.; Housiaux, P.; Adams, J.M.

EMBO J. 7, 2131-2138, 1988

A;Title: Expression of multiple homeobox genes within diverse mammalian haemopoietic lin

A;Reference number: S00987; MUID:88329001; PMID:2901346

A;Accession: S00988

A;Molecule type: mRNA

A;Residues: 137-196 <KON>

A;Cross-references: EMBL:X14570; NID:G51388; PIDN:CAA32708.1; PID:G930147

R;Verrilizer, P.; de Graaf, W.; Deschamps, J.; Weilink, F.

Nucleic Acids Res. 16, 2729, 1988

A;Title: Nucleotide sequence of the Hox2.3 gene region.

A;Reference number: I48411; MUID:88203221; PMID:2896332

A;Accession: I48411

A;Status: translation not shown; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-217 <RES>

A;Cross-references: EMBL:X06762; NID:G51389; PIDN:CAA29934.1; PID:G51390

C;Genetics:

A;Gene: Hox-2.3

A;Map position: 11

A;Superfamily: homeotic protein Hox A7; homeobox homology

C;Cross-references: homeobox: nucleus: transcription regulation

C;Accession: A28329; I49131
R;Kessel, M.; Schulze, F.; Fibi, M.; Gruss, P.
Proc. Natl. Acad. Sci. U.S.A. 84, 5306-5310, 1987
A;Title: Primary structure and nuclear localization of a murine homeodomain protein.
A;Reference number: A28329; MUID:87260976; PMID:2885847
A;Accession: A28329
A;Molecule type: mRNA
A;Residues: 1-229 <RES>
A;Cross-references: UNIPROT:P02830; GB:M17192; NID:913906; PIDN:AAA37833.1; PID:g309313
R;Parikh, H.; Shah, S.; Hilt, D.; Peterkofsky, A.
Gene 154, 237-242, 1995
A;Title: Organization, sequence and regulation of expression of the murine Hoxa-7 gene.
A;Reference number: I49131; MUID:95197009; PMID:7890170
A;Accession: I49131
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-42, 'GAG', 46-229 <RES>
A;Cross-references: EMBL:U15972; NID:g664757; PIDN:AAC52160.1; PID:g664758
C;Genetics:
A;Gene: Hoxa7
A;Introns: 126/1
C;Superfamily: homeotic protein Hox A7; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
P;130-186/Domain: homeobox homology <HOX>

Query Match 14.4%; Score 201.5; DB 2; Length 229;
Best Local Similarity 26.3%; Pred. No. 2e-09;
Matches 68; Conservative 33; Mismatches 71; Indels 87; Gaps 12;

Qy 19 FSEKAPMCQV--SVNSFYKLSPGLPQLAAGTGHGTTILSRPVATPNSLLSGYPH 76
Db 21 FQNAEPTSCSPAPNSQSGY-----GPAPAFASTVPLGLYN-----NSPLYQS-PF 66

Qy 77 VAGFG-----GLSS-----QGVYGPQVGSFSGKAGNEYPTR 107
Db 67 ASGYGLGADAYNLPCASYDQNIPLGLCSLDLAKGACDAEGVLHGPAEASF-----RIYP-- 120

Qy 108 TRNCWADTGDWGRSARPCGNTPDPLSDTIHKKHTRPTFTGHQIFALEKTFEOTKYLAG 167
Db 121 -----WMRSSGP-----DRKRGQTYTRYQTLEKEFHFNFYLTR 156

Qy 168 PERARLAYSLGWTESQVQVWFQNRRTKWRKSALEPSSSTPRAPCGA-----SGDRAAS 221
Db 157 RRRIEHALCLUTERQIKIWFQNRKMKKEKHDESQAPT-AAPEDAVPSVSTADKADE 215

Qy 222 ENEDDYNNKPLDPDSDDEK 240
Db 216 EEEEE-----EEEEEE 229

RESULT 24
T22698
hypothetical protein F55B12.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 16-Aug-2004
A;Accession: T22698
R;Sims, M.
submitted to the EMBL Data Library, September 1996
A;Reference number: Z19602
A;Accession: T22698
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-232 <WIL>
A;Cross-references: UNIPROT:Q9NLC2; EMBL:279757; PIDN:CAB02124.1; GSPDB:GN000023; CESP:F55B12
A;Experimental source: clone F55B12
C;Genetics:
A;Gene: CESP:F55B12.1
A;Map position: 5
A;Introns: 37/2; 126/3; 168/2; 197/3
C;Superfamily: homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match 14.4%; Score 201; DB 2; Length 232;

Best Local Similarity 30.5%; Pred. No. 2.2e-09;
Matches 58; Conservative 36; Mismatches 66; Indels 30; Gaps 8;

Qy 65 TPNSLLSGYPHVAAGFGLSSQGVY--GPQVGSFSGKAGNEYPTRTRNCWADTGDWGRSA 123
Db 8 TPGTN--AGVGAFFPYGPGRLPGNYFAGPPPGYSGAQPWY-----NGNDPRFAA 55

Qy 124 R-----PCGNTPDPLSDTIH-----KKHTRPTFTGHQIFALEKTFEOTKYLAGPER 170
Db 56 AAALLPC--SIDPVSAINHQFSMSMSQRRKRVLFSAQVYELERRFKQAKYLTAPEP 113

Qy 171 ARLAYSLGWTESQVQVWFQNRRTKWRKSALEPSSSTPRAPCGASGDRAAASENEDDYNNK 230
Db 114 EQLANSIRLPTPTQKIWFQNRHYCKRQEKAKMSGLGHSDEGSS--PPPDNDDDDDKYSI 172

Qy 231 PLDPDSDDEK 240
Db 173 EMD-DKDEE 181

RESULT 25
I50112
Dlx4 homeodomain protein - zebra fish
N;Alternate names: distal-less X4 protein
C;Species: Brachydanio rerio (zebra fish)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 16-Aug-2004
A;Accession: I50112
R;Akimenko, M.A.; Ekker, M.; Wegner, J.; Lin, W.; Westerfield, M.
J. Neurosci. 14, 3475-3486, 1994
A;Title: Combinatorial expression of three zebrafish genes related to distal-less: part 1
A;Reference number: I50111; MUID:94267510; PMID:7911517
A;Accession: I50112
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-283 <AKI>
A;Cross-references: UNIPROT:P50576; EMBL:U03876; NID:g460128; PIDN:AAA19827.1; PID:g460128
C;Genetics:
A;Gene: dlx4
C;Superfamily: homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
P;138-194/Domain: homeobox homology <HOX>

Query Match 14.3%; Score 200; DB 2; Length 283;
Best Local Similarity 32.8%; Pred. No. 3.4e-09;
Matches 57; Conservative 23; Mismatches 74; Indels 20; Gaps 4;

Qy 48 AAGTPHGTTILSRPVATPNSLLSGYPHVAAGFGLSSQGVYGPQVGSFSGKAGNEYPTR 107
Db 50 AGVHHGYCSPNSGTGKPLNAYQYHGVNGSSGNYSAKSY--PDYGSYSTAYHQYAGT 107

Qy 108 TRNCWADTGDWGRSARP-----CGNTPDPLSDTIHKKHTRPTFTGHQIFALEKTFEOTK 163
Db 108 YNRVQSQSPSQEKETAEPVRMVGKP-----KKVRKPRTIYSSFFQLAALQRRFQNTQ 160

Qy 164 YLACPERARLAYSLGWTESQVQVWFQNRRTKWRK-----KSALEPSSSTPRA 210
Db 161 YLALPERARLAASLGLTQTQKIWFQNRKSKLKKIMKNGELPPEHSPSSSDPMA 214

RESULT 26
I48690
homeotic protein NKx-5.1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Aug-2004
A;Accession: I48690; S37562
R;Bober, E.; Baum, C.; Braun, T.; Arnold, H.H.
Dev. Biol. 162, 288-303, 1994
A;Title: A novel NK-related mouse homeobox gene: expression in central and peripheral ne
A;Reference number: I48690; MUID:94170918; PMID:7510254
A;Accession: I48690
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-458 <RES>

Db 203 GRPAGVHLGLOPSAGQFFASLDPTSEASAILSPSSNPRNSVQHOFD-----TF 253
 Qy 130 PDPLS-----DTI-----HKKHTRPTFTGHQIFALEKTFEOTKYLAGEPARARLAYSLGMT 180
 Db 254 PGPAVLTKDTPQTKYKRSWRAVFNLOKRGLEKFEIQTQYVTRKPDRLKOLAAIIGLT 313
 Qy 181 ESQVKVWFONRTKWR-----KKSALPSSSTP-----RAPGASGDRAAS 221
 Db 314 DAQVKVWFONRMKWRHSEKQAQKDKKEAGKPSGVPAGEBREERSPSRSEGE-AES 372
 Qy 222 ENEDDEYNKPLD-PDSDDK 240
 Db 373 ESSDSE---SLDAPSDTER 389

RESULT 33
 A61045
 homeotic protein Tghbox 5 - sea urchin (Tripneustes gratilla) (fragment)
 C:Species: Tripneustes gratilla
 C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
 C:Accession: A61045
 R:Wang, G.V.L.; Dolecki, G.J.; Carlos, R.; Humphreys, T.
 Dev. Genet. 11, 77-87, 1990
 A:Title: Characterization and expression of two sea urchin homeobox gene sequences.
 A:Reference number: A61045; MUID:90298585; PMID:1972915
 A:Accession: A61045
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-232 <WAN>
 A:Cross-references: UNIPROT:Q7M3U6
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:81-137/Domain: homeobox homology <HOX>

Query Match 14.0%; Score 196; DB 2; Length 232;
 Best Local Similarity 37.2%; Pred. No. 5.7e-09;
 Matches 54; Conservative 18; Mismatches 57; Indels 16; Gaps 4;

Qy 62 PVATP--NSSLLSGYPHVGAGLSSQGVYYPQVGSFKAGNEY--PTRTRNCWADTGO 117
 Db 10 PIAQPLTNSSRLAMAKDVSPRTQSHS--FGESGEKDSRDSPDELRNDGHEEE 67
 Qy 118 DWRGARPCGNTPDPLSDTIHKHTRPTFTGHQIFALEKTFEOTKYLAGEPARARLAYSL 177
 Db 68 D-----DDDRSPQKKKKKTRTVFSRSQVFSQVLESFVVKRYLSSSERAGLAANL 117
 Qy 178 GMTESQVKVWFONRTKWRKKSAL 202
 Db 118 HLTETQVKIWFQNRNKKRWMAAE 142

RESULT 34
 S41860
 gene Nkx-1.1 protein - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 16-Aug-2004
 C:Accession: S41860
 R:Schubert, F.R.; Gruss, P.
 submitted to the EMBL Data Library, October 1993
 A:Description: Expression of the novel Murine homeobox gene Nkx-1.1 in the developing ne
 A:Reference number: S41860
 A:Accession: S41860
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-305 <SCH>
 A:Cross-references: UNIPROT:P42580; EMBL:X75384; NID:9453171; PIDN:CAAS5153.1; PID:94531
 C:Superfamily: homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:157-213/Domain: homeobox homology <HOX>

Query Match 14.0%; Score 196; DB 2; Length 305;
 Best Local Similarity 36.2%; Pred. No. 8e-09;

Matches 55; Conservative 17; Mismatches 52; Indels 28; Gaps 7;
 Qy 86 QGVVYG-PQ-----VGSFGKAGNEYPTTRNCWADTQDWRGSRAPCGNTPDPLSDTIHK 139
 Db 108 QGVHESPEARAVAVGTEESGAGLP-----ASPGSP--GSPRRRRAE---SSCAK 155
 Qy 140 KKHTRPTFTGHQIFALEKTFEOTKYLAGEPARARLAYSLGMTESQVKVWFONRTKWRKKS 199
 Db 156 PRARTAFTYEQVLVALENKFRATRYLSVCERLNLALSLSLTETQVKIWFONRTKWKON 215
 Qy 200 -----ALE-----PSSSTPRAPGASGDRAAS 221
 Db 216 PGADGAVOAGGAPQPGTTCAGVAGGSGSATGS 247

RESULT 35
 JC6307
 homeobox protein Barx2 - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Aug-2004
 C:Accession: JC6307
 R:Jones, F.S.; Kloussi, C.; Copertino, D.W.; Kallunki, P.; Holst, B.D.; Edelman, G.M.
 Proc. Natl. Acad. Sci. U.S.A. 94, 2632-2637, 1997
 A:Title: Barx2, a new homeobox gene of the Bar class, is expressed in neural and craniofac
 A:Reference number: JC6307; MUID:97226005; PMID:9122247
 A:Accession: JC6307
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-258 <JON>
 A:Cross-references: UNIPROT:Q921G1; GB:L77900
 C:Superfamily: homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match 14.0%; Score 195.5; DB 2; Length 258;
 Best Local Similarity 24.5%; Pred. No. 7.1e-09;
 Matches 60; Conservative 35; Mismatches 77; Indels 73; Gaps 7;

Qy 27 COYVSQNSFYKLSPPGL-----GPQLAA-----GTPHGITD 57
 Db 11 CDYPEKLSLYSVCPSLVVRPKPLHSCGSPSLRAYPLPSVITRQPTVISHLVPTGSLTP 70
 Qy 58 ILSR-PVATPNSSLLSGYPHVGAGLSSQGVYYPQVGSFKAGNEYPTTRNCWADTG 116
 Db 71 VLTRHPVAAAEAAAAA--ETPGGEALAS-----SESETEQPTPRK----- 110
 Qy 117 QDMRGARPCGNTPDPLSDTIHKHTRPTFTGHQIFALEKTFEOTKYLAGEPARARLAY 176
 Db 111 -----KPRSRKTIFTLQLMGMEKKFKQKYLSTPDRDLAQS 148
 Qy 177 LGMTESQVKVWFONRTKWRKKSALPSSSTPRAPGASGDRAASENEDDEYNKPLDPDS 236
 Db 149 LGLTQLQVKTWYQNRMRK-KQVYLGQGEAPTQKGRPKKNISPTSEEIEAEKMNQSA 207
 Qy 237 DDEKI 241
 Db 208 SQSEL 212

RESULT 36
 C88429
 protein ceh-43 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 16-Aug-2004
 C:Accession: C88429
 R:anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolo
 A:Reference number: A75000; MUID:99069613; PMID:9851916
 A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_ele
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A:Accession: C88429
 A:Status: preliminary
 A:Molecule type: DNA

A;Residues: 1-273 <STO>
A;Cross-references: UNIPROT:Q18273; GB:chr_III; PIDN:CAA83601.1; PID:g3874518; GSPDB:GNC
C;Genetics:
A;Gene: ceh-43
A;Map position: 3
C;Superfamily: homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match 14.0%; Score 195.5; DB 2; Length 273;
Best Local Similarity 29.7%; Pred. No. 7.6e-09;
Matches 58; Conservative 31; Mismatches 57; Indels 49; Gaps 8;

Qy 73 GPEVAGGGLSSQGVYGPQVGSFKAG-----NEYPTRTNCWADTGQDWGSGARPC- 126
Db 6 GFEYVAG-DYQTSGVAPPTNSGAGNSVPFYPHYPTSTN--GATGGSGYGPQOTS 62

Qy 127 -----GNTPD---PLSDTI-----HKKGHTRPTFTGHQIFALEKTFE 160
Db 63 AVAMYPGPGSSPEAFPEHITTKIVEGCEAKYNVKGKMKRPRTIYNSQLQMKKFKQ 122

Qy 161 QTKYLAGEPARLAYSLGWTESQVKWFQNRRTKWKKSALPPSSSTPRAPGSGDRAA 220
Db 123 KTOYLALDPRALAHGLSLQTVQKWFQNRRTKWKKSALPPSSSTPRAPGSGDRAA 220

Qy 221 SENEDD-EYNKPLDP 234
Db 170 DEEDDDTESKPESP 184

RESULT 37
A36664
S59/2 homeotic protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 19-Apr-1991 #sequence revision 19-Apr-1991 #text_change 09-Jul-2004
C;Accession: A36664; B36664; A33976; S14404
R;Dohmann, C.; Azpiazu, N.; Frasch, M.
Genes Dev. 4, 2098-2111, 1990
A;Title: A new Drosophila homeo box gene is expressed in mesodermal precursor cells of
A;Reference number: A36664; MUID:91099659; PMID:1980118
A;Accession: A36664
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-659 <DOH>
A;Cross-references: UNIPROT:P22807; GB:X55393; NID:98530; PIDN:CAA39067.1; PID:98531
A;Accession: B36664
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-123 'S', 125-144, 'M', 146-204, 'QH', 207-335, 'T', 337-659 <DO2>
A;Cross-references: GB:X55393
R;Kim, Y.; Nirenberg, M.
Proc. Natl. Acad. Sci. U.S.A. 86, 7716-7720, 1989
A;Title: Drosophila NK-homeobox genes.
A;Reference number: A33976; MUID:90046666; PMID:2573058
A;Accession: A33976
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 497-625 <KIM>
A;Cross-references: GB:M27289; NID:9157633; PIDN:AAA28616.1; PID:g157634
C;Genetics:
A;Gene: FlyBase:Nkl
A;Cross-references: FlyBase:FBgn0002941
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:546-602/Domain: homeobox homology <HOX>

Query Match 14.0%; Score 195.5; DB 2; Length 659;
Best Local Similarity 40.6%; Pred. No. 2.3e-08;
Matches 43; Conservative 16; Mismatches 26; Indels 21; Gaps 2;

Qy 131 DPLSDTIHKK-----KHTRPTFTGHQIFALEKTFEQTLYLAGPER 170
Db 516 DPSETDSKGGSRNGDKSGGGGGGSKPRPRTAFTYEQVLVLENKFKTRYLSVCR 575

A;Residues: 1-273 <STO>
A;Cross-references: UNIPROT:Q18273; GB:chr_III; PIDN:CAA83601.1; PID:g3874518; GSPDB:GNC
C;Genetics:
A;Gene: ceh-43
A;Map position: 3
C;Superfamily: homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match 14.0%; Score 195; DB 2; Length 220;
Best Local Similarity 29.2%; Pred. No. 6.4e-09;
Matches 69; Conservative 28; Mismatches 69; Indels 70; Gaps 12;

Qy 19 FSEMKAPCQVSVQNSFYKLSPPGLGQLAAGTGHGIDILSRPVATNSLLSGYPHVA 78
Db 26 FSEQTSCAFASPPQSGYNSPGGTFF--AGSAAHGLFS-----NGSSL--HP--- 69

Qy 79 GFGLSSQGVY---YQGVGSFSKAGNEY-----PTRTRNCWADTGQD----- 118
Db 70 -----QGPAMYPSYGLDAASFMHCHSPFEONLSLMCDPTK-QNCTKAEQDSSELHNEA 123

Qy 119 -----NRGSARPCGNTPDPLSDTIHKKHTRPTFTGHQIFALEKTFEQTLYLAGPERAR 172
Db 124 NLRIYPMWSAG-----SD-----RKRCQYTRYQTLELEKEFHFNRYLTRRRIE 170

Qy 173 LAYSLGWTESQVKWFQNRRTKWKKS-ALPSSSTPRAPGSGDRASNNEDDE 227
Db 171 IAHTLCITERQIKIWFQNRRTKWKKS-ALPSSSTPRAPGSGDRASNNEDDE 227

RESULT 39
S60249
homeotic protein flh - zebra fish
C;Species: Brachydanio rerio (zebra fish)
C;Date: 10-Apr-1996 #sequence revision 19-Apr-1996 #text_change 16-Aug-2004
C;Accession: S60249
R;Talbot, W.S.; Trevarrow, B.; Halpern, M.E.; Melby, A.E.; Farr, G.; Postlethwait, J.H.,
Nature 378, 150-157, 1995
A;Title: A homeobox gene essential for zebrafish notochord development.
A;Reference number: S60249; MUID:96069732; PMID:7477317
A;Accession: S60249
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-241 <TAL>
A;Cross-references: UNIPROT:Q90461; EMBL:L48017; NID:g1063603; PIDN:AAC42230.1; PID:g101
C;Superfamily: homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:135-191/Domain: homeobox homology <HOX>

Query Match 14.0%; Score 195; DB 2; Length 241;
Best Local Similarity 38.1%; Pred. No. 7.2e-09;
Matches 43; Conservative 21; Mismatches 35; Indels 14; Gaps 3;

QY 139 KKKHTRPTFTGHOIFALEKTFEQKYLGPAGPERARLAYSLGWTESQVKVWFONRRTKWK 198
DB 133 KSKNRMTSFTNDQLSRLEKEFARQQYMWGSEFLLASALQLEAQVKVWFONRIKWK 192
QY 199 SALEPSSS-----TPRAPGASGDRAASENEDDEYNKLPDPSD-DEKIR 242
DB 193 SLEQQQAKLAKLGLTVPKPSGS-----QGREDEERDFTESDVIDIDDSLQ 240

RESULT 40
S42634
homeoprotein, IDX-1 - rat
N:Alternate names: somatostatin transactivating factor-1
C:Species: Rattus norvegicus (Norway rat)
C:Date: 25-Dec-1994 #sequence_revision 03-Aug-1995 #text_change 16-Aug-2004
C:Accession: S42634; A47750
R:Miller, C.P.; McGehee Jr., R.E.; Habener, J.F.
EMBO J. 13, 1145-1156, 1994
A:Title: IDX-1: a new homeodomain transcription factor expressed in rat pancreatic islet
A:Reference number: S42634; MUID:94178257; PMID:7907546
A:Accession: S42634
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-283 <MIL>
A:Cross-references: UNIPROT:P52947; GB:U04833; NID:G454391; PID:AAA18355.1; PID:G454392
R:Leonard, J.; Peers, B.; Johnson, T.; Ferreri, K.; Lee, S.; Montminy, M.R.
Mol. Endocrinol. 7, 1275-1283, 1993
A:Title: Characterization of somatostatin transactivating factor-1, a novel homeobox fac
A:Reference number: A47750; MUID:94088595; PMID:7505393
A:Accession: A47750
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-46, 'P', 48-239, 'P', 241-246, 'A', 248-283 <LEO>
A:Cross-references: GB:S67435; NID:G456961; PID:AA829317.1; PID:G456962
A:Experimental source: TU-6 cells
A>Note: sequence extracted from NCBI backbone (NCBIN:141266, NCBIP:141267)
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:147-203/Domain: homeobox homology <HOX>

Query Match 14.0%; Score 195; DB 2; Length 283;
Best Local Similarity 29.9%; Pred. No. 8.8e-09;
Matches 66; Conservative 23; Mismatches 86; Indels 46; Gaps 11;

QY 25 PMCOYSQNSFYKLSPPGLGPOLAA-----GTPHGITDILSRPVA-TPNSSLLSGYPH 76
DB 33 PACLY-----MGQPPPTPTPQFAGSLGLEQSPDIPSYEVPLADDPAGAH--HHH 85
QY 77 VAGFGGLSSQGVYGVQVGSFKAGNEYPTTR-----NCWADTGODWEGSARP 125
DB 86 LPAQLGLAHP--GFPNGTGTGLEPSRVHLPPPMKSTKAHAWK---SQWAGGA-- 138
QY 126 CGNTPPLSDTIHKKHPTFTGHOIFALEKTFEQKYLGPAGPERARLAYSLGWTESQVK 185
DB 139 --YAAEF-----EENKRTTAYTQALLELEKEFLFNKYISRRVRELAVMLNTERHK 191
QY 186 VWFQNRRTKWKKSALPSSSTPRAPGASGDRAASENEDD 226
DB 192 INFQNRMKKKEEDKRSST--TSGGGGGE---EPQD 226

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13., 2005, 17:01:49 ; Search time 252.264 Seconds
(without alignments)
531.844 Million cell updates/sec

Title: US-09-998-861-13
Perfect score: 1396
Sequence: 1 MESNLQTFLLNVTQAQFS.....LLLRKHRAAFVSLGHSV 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1385	99.2	262	Q9D2W8	Q9D2W8 mus musculus
2	590	42.3	277	Q9C056	Q9C056 homo sapien
3	565	40.5	312	Q6DGKS	Q6DGKS brachydanio
4	558.5	40.0	365	1_NK61_MOUSE	Q99ma9 mus musculus
5	558.5	40.0	365	1_NK61_RAT	Q35762 rattus norv
6	557	39.9	367	1_NK61_HUMAN	P78426 homo sapien
7	553	39.6	364	1_NK61_MESAU	Q60554 mesocricetu
8	430.5	30.8	153	2_Q9YHC2	Q9Yhc2 gallus gall
9	381	27.3	135	2_Q96LR0	Q96lr0 homo sapien
10	347	24.9	480	2_Q8SXX1	Q8sxx1 drosophila
11	347	24.9	513	2_Q9NHP8	Q9nhp8 drosophila
12	331	23.7	108	2_Q7OC36	Q7oc36 anopheles g
13	308	22.1	106	2_Q8G1L7	Q8g1l7 sus scrofa
14	267	19.1	106	2_Q81LK2	Q81lk2 caenorhabdi
15	267	19.1	256	2_Q17978	Q17978 caenorhabdi
16	239.5	17.2	214	2_Q8CF07	Q8cf07 mus musculus
17	239	17.1	380	2_Q6E7C7	Q6e7c7 oikopleura
18	228	16.3	277	1_HMPH_CHICK	Q05502 gallus gall
19	227	16.3	284	1_TLX2_HUMAN	Q43763 homo sapien
20	223	16.0	297	2_Q17319	Q17319 caenorhabdi
21	222	15.9	309	2_Q6SYZ0	Q6syz0 brachydanio
22	221.5	15.9	284	1_TLX2_MOUSE	Q61663 mus musculus
23	221.5	15.9	320	2_Q6SVY9	Q6svy9 brachydanio
24	220	15.8	641	2_Q61640	Q61640 drosophila
25	220	15.8	721	2_Q9VFK4	Q9vfk4 drosophila
26	219.5	15.7	322	2_Q7KVFA	Q7kvfa drosophila
27	219.5	15.7	327	1_HMDL_DROME	P20009 drosophila
28	219.5	15.7	327	2_Q9W0Z8	Q9w0z8 drosophila
29	218	15.6	132	2_Q9GP46	Q9gp46 diacocelis
30	218	15.6	333	2_Q9QX99	Q9qx99 mus musculus
31	217.5	15.6	271	1_HMPH_MOUSE	P43120 mus musculus

32	217.5	15.6	280	2_Q9CRV1	Q9crv1 mus musculus
33	217.5	15.6	292	2_Q9DFC2	Q9dfc2 xenopus lae
34	217.5	15.6	297	1_TLX3_CHICK	Q93367 gallus gall
35	217	15.5	284	1_IPF1_MOUSE	P52946 mus musculus
36	216	15.5	271	2_Q9R1X2	Q9r1x2 mus musculus
37	215.5	15.4	332	1_TLX1_MOUSE	P43345 mus musculus
38	214.5	15.4	388	2_Q9W7I0	Q9w7i0 pleurodeles
39	214.5	15.4	857	1_PAX6_DROME	Q18381 drosophila
40	212.5	15.2	328	1_DLK2_HUMAN	Q07687 homo sapien
41	212	15.2	291	1_TLX3_HUMAN	Q43711 homo sapien
42	212	15.2	291	1_TLX3_MOUSE	Q55144 mus musculus
43	212	15.2	400	2_Q9DDF9	Q9ddf9 petromyzon
44	212	15.2	405	1_HMX_STRPU	Q26656 strongyloce
45	211.5	15.2	275	2_Q8MZL4	Q8mzl4 branchiosteo

ALIGNMENTS

RESULT 1
Q9D2W8 PRELIMINARY; PRT; 262 AA.
ID Q9D2W8;
AC Q9D2W8; (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 26, Last annotation update)
DE Mus musculus adult male cecum cDNA, RIKEN full-length enriched
DE library, clone:9130417107 product:hypothetical Homeobox domain
DE containing protein, full insert sequence.
GN Name=9130417107Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cecum;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cecum;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN PANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cecum;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cecum;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,

RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cecum;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayashizaki N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Komori H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numasaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai K., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Teijima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR EMBL; AK018683; BAB31342.1; -.
DR HSSP; PI3297; 1IG7.
DR MGD; MG1:192181; 9130417107Rik.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeobox-like.
DR InterPro; IPR000047; HTH_lambdrepresr.
DR Pfam; PF00046; Homeobox-1.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00031; HTHREPRESSR.
DR PRODOM; PD00010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS50071; HOMEBOX 2; 1.
DR PROSITE; PS50071; HOMEBOX 2; 1.
KW DNA-binding; Homeobox; Hypochemical protein; Nuclear protein.
SQ SEQUENCE 262 AA; 28778 MW; B4554F3D60DF10B6 CRC64;

Query Match 99.2%; Score 1385; DB 2; Length 262;
Best Local Similarity 99.2%; Pred. No. 3.3e-106;
Matches 260; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MESNLQGTFLNNTQLAQSSEKAPMCQSVQNSFYKLSPPGLGQLAAGTGHGTDILS 60
DB 1 MESNLQGTFLNNTQLAQSSEKAPMCQSVQNSFYKLSPPGLGQLAAGTGHGTDILS 60

QY 61 RPVATNSLLSGYPHVAGFGLSSQGVYGVQVGSFSGAGNEYPTRTNRCWADTQDWR 120
DB 61 RPVATNSLLSGYPHVAGFGLSSQGVYGVQVGSFSGAGNEYPTRTNRCWADTQDWR 120

QY 121 GSARPCGNTDPDLSDTIHKKKHTRPTTGHQIPALEKTEQTKYLAGPERARLAYS LGMT 180
DB 121 GSARPCGNTDPDLSDTIHKKKHTRPTTGHQIPALEKTEQTKYLAGPERARLAYS LGMT 180

QY 181 ESQVKVWFQNRRTKWKRSALPSSSTPRAPGASGDRAASENEDDEYNKPLDPSDDDK 240
DB 181 ESQVKVWFQNRRTKWKRSALPSSSTPRAPGASGDRAASENEDDEYNKPLDPSDDDK 240

QY 241 IRLLLKRAAFSVLSLGAHSV 262
DB 241 IRLLLKRAAFSVLSLGAHSV 262

RESULT 2
Q9C056 PRELIMINARY; PRT; 277 AA.
AC Q9C056; 2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Name=GTx;
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21077538; PubMed=11210186;
RA Lee S.H., Davison J.A., Vidal S.M., Belouchi A.,
RT "Cloning, expression and chromosomal location of NKX6B to 10Q26, a
RT region frequently deleted in brain tumors.";
RL Mamm. Genome 12:157-162(2001).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR EMBL; AF184215; AAK13251.1; -.
DR HSSP; PI3297; 1IG7.
DR TRANSFAC; T04765; -.
DR TRANSFAC; T04787; -.
DR Genew; HGNC:19321; NKX6-2.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0003700; F:transcription factor activity; NAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeobox-like.
DR InterPro; IPR000047; HTH_lambdrepresr.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00031; HTHREPRESSR.
DR PRODOM; PD00010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS50071; HOMEBOX 2; 1.
DR PROSITE; PS50071; HOMEBOX 2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 277 AA; 29235 MW; 4205E50F33AC4B6 CRC64;

Query Match 42.3%; Score 590; DB 2; Length 277;
Best Local Similarity 50.8%; Pred. No. 1.5e-40;
Matches 133; Conservative 28; Mismatches 85; Indels 16; Gaps 8;

QY 1 MESNLQGTFLNNTQLA---QFSKAPMCQSVQNSFYKLSPPGLGQLAAGTGHG 55
DB 1 MDTRPGAFVLSAPLAALHNMAEMKTSLPYALQGPAGFKAPALGGLGQLPLGTGPHG 60

QY 56 TDILSRPVATNSLLSGYPHVAGFGLSSQGVYGVQVGSFSGAGNEYPTRTNRCW 112
DB 61 SDILGRPVGAAGGLGGLPRLNGLA--SSAGVYVGPAAAVARGYPKPLAELPGRPIFW 118

QY 113 ADT--GQWRGSRPCGNTP--DPLSDTIHKKKHTRPTTGHQIPALEKTEQTKYLAGPE 169
DB 119 PGVYQGPWR--DPLAGAPAGGVLDKDGKKHSRPTFSGQIIPALEKTEQTKYLAGPE 177

QY 170 RARLAYS LGMTESQVKVWFQNRRTKWKRSALPSSSTPRAPGASGDRAASEN--EDDE 227
DB 178 RARLAYS LGMTESQVKVWFQNRRTKWKRSALPSSSTPRAPGASGDRAASEN--EDDE 227

QY 228 YNKPLDPSDDDKIRLLLRKH 249
DB 238 YNRPLDPSDDDKIRLLLRKH 259

RESULT 3
Q6DGK5 PRELIMINARY; PRT; 312 AA.
AC Q6DGK5; 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Zgc:92879.
GN Name=zgc:92879;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostarlophysi; Cypriniformes;
OC Cyprinidae; Danio.

OX NCBI_TaxID=7955;
 RN
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M.J., Udwin T.B., Toshitake S., Carninci P., Prange C.,
 RA Brownstein M.J., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 DR EMBL; BC076337; AAH76337.1; -;
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR009057; Homeobox-like.
 DR InterPro; IPR000047; HTH_lambdarepressr.
 DR Pfam; PF00046; Homeobox; 1.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR PRINTS; PR00024; HOMEBOX.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS00071; HOMEBOX_2; 1.
 DR DNA-binding; Homeobox; Nuclear protein.
 SQ SEQUENCE 312 AA; 33801 MW; 019422BADD58C5B2 CRC64;

 Query Match 40.5%; Score 565; DB 2; Length 312;
 Best Local Similarity 46.0%; Pred. No. 28-38;
 Matches 134; Conservative 25; Mismatches 90; Indels 42; Gaps 8;

 QY 1 MESNLQGTFLNNNTOLA---QFSEMKAPM-----CQSVQNSFYKLSPP 40
 DB 7 MDGSRQSAFLNTPPLAALHSMWEMTKTPLYPAYLSSTGPASTSTPATSPPGIPVSS 66

 QY 41 PGL-----GQLAAGTHGTDILSRP-VATPNSSLLSGYVHVGGLSSQGV 88
 DB 67 PGKTSSTGLSALASAOQAIAIAPHGINDILSRPVSACSPAGILSLPRPSSLPPLPGL 126

 QY 89 YVGPO-----VGSFSGAGNEYTRTNCWADTGQD--WRGSARPCG-NTPDPLSDTIHKK 140
 DB 127 YFSPSAAAVAVARYPKPLTELPGRTPIFWPVGVMQSPHWRDARFACSPHNSVLLDKGKR 186

 QY 141 KTRTPFTTGHQIFALEKTEQTKYLAGPERARLAYSLGNTESQVKKVFNRRRTKWKKSA 200
 DB 187 KTRTPFTSQQIIFALEKTEQTKYLAGPERARLAYSLGNTESQVKKVFNRRRTKWKKHA 246

 QY 201 LEPSTSTPPRAGSGADRAASENE--DDEYNKPLDPDSDEKIRLLLRKH 249
 DB 247 AEMASAKKKQDSETERLKCASENEDDDDDYNNKPLDPNSDDEKITQLLKKHK 297

RESULT 4
 NK61_MOUSE
 ID NK61_MOUSE STANDARD; PRT; 365 AA.
 AC Q99MA9; Q9ERQ7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Homeobox protein Nkx-6.1.
 OS Name=Nkx6.1; Synonyms=Nkx6.1, Nkx6a;
 GN Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 X DBA/2;
 RA Sander M., Nelson S.B.;
 RT "Isolation of partial cDNA sequence for the mouse homeodomain protein,
 RT Nkx6.1";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-224 FROM N.A.
 RX MEDLINE=20507862; PubMed=10938085; DOI=10.1074/jbc.M004981200;
 RA Watada H., Mirmira R.G., Leung J., German M.S.;
 RT "Transcriptional and translational regulation of beta-cell
 RT differentiation factor Nkx6.1";
 RL J. Biol. Chem. 275:34224-34230 (2000).
 CC -1- FUNCTION: May be important for control of islet development and/or
 CC regulation of insulin biosynthesis (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: Contains 1 homeobox domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AF357883; AAK37567.1; -;
 DR EMBL; AF291666; AAG30415.1; -;
 DR HSSP; P13297; 1IG7.
 DR TRANSFAC; T04269; -;
 DR MGD; MGI:1206039; Nkx6-1.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR009057; Homeobox-like.
 DR InterPro; IPR000047; HTH_lambdarepressr.
 DR Pfam; PF00046; Homeobox; 1.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR PRINTS; PR00024; HOMEBOX.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS00071; HOMEBOX_2; 1.
 DR Developmental protein; DNA-binding; Homeobox; Nuclear protein.
 KW DOMAIN 49 61 Poly-Ser.
 FT DOMAIN 119 132 Poly-Ser.
 FT DOMAIN 136 151 Poly-Ala.
 FT DOMAIN 169 174 Poly-Pro.
 FT DNA BIND 237 296 Homeobox.
 FT DOMAIN 323 326 Poly-Asp.
 SQ SEQUENCE 365 AA; 37705 MW; B6AABAF5748A99E8 CRC64;

 Query Match 40.0%; Score 558.5; DB 1; Length 365;
 Best Local Similarity 41.5%; Pred. No. 8.4e-38;
 Matches 148; Conservative 26; Mismatches 80; Indels 103; Gaps 13;

 QY 1 MESNLQGTFLNNNTOLA---QFSEMKAPM-----CQSVQNSFYKLSPP--- 41
 DB 7 MEGPRQSAFLSSPPPLAALHSMWEMTKTPLYPAAYPLPTGTPSSSSSSSSSPPLGCS 66

QY 42 -----GLGP-----OLAACTPHGTDILSRP-----VATPNS-- 69
 DB 67 HNPGLKPPAAGSLGSLGPPQOLSAATPHGINDILSRPMPVAGSALPSGSSSS 126
 QY 70 -----GLGP-----LLSGYPHVAGFGGLS-----SQGVYGPQ-- 93
 DB 127 SSSASATSASAAAAAASAPAGLLAGLPR---FSSLSPPPPPGLYFSPSAA 183
 QY 94 -----VGSFSGKAGNEYPTTRNCWADTGD--WRGSARPCGNTDPD---LSDTIHKKHTR 144
 DB 184 AVAAGVRYPKPLAELPGRTPIFWPGVQSPWDRALAC--TPHQGSILLDKDKKHKTR 241
 QY 145 PTFTGHOIFALEKTFEOTKYLAGPERARLAYSLGWTESQVQVWFQNRRTKWKKSALPS 204
 DB 242 PTFSGQQIFALEKTFEOTKYLAGPERARLAYSLGWTESQVQVWFQNRRTKWKKHAEMA 301
 QY 205 SSTPRAPGASGDRAASENE--DDYNKPLDPDSDEKIRILLKRAAFSVLSGA 259
 DB 302 TAKKKQDSETERLKGTSENEEDDDYNKPLDPNSDEKITQLLKKHSGGSLHLHA 358
 RESULT 5
 NK61_HUMAN
 ID NK61_HUMAN STANDARD; PRT; 365 AA.
 AC Q35762;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Homeobox protein Nkx-6.1.
 GN Name=Nkx6-1; Synonyms=Nkx6.1, Nkx6a;
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus;
 OC NCBI_TaxID=10116;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=WiStar; TISSUE=Pancratic islets;
 RX MEDLINE=20036461; PubMed=10567713; DOI=10.1016/S0014-5793(99)01436-2;
 RA Jorgensen M.C., Vestergaard Petersen H., Ericson J., Madsen O.D.,
 Serup P.,
 RT "Cloning and DNA-binding properties of the rat pancreatic beta-cell-
 specific factor Nkx6.1";
 RL FEBS Lett. 461:287-294 (1999).
 CC -!- FUNCTION: May be important for control of islet development and/or
 regulation of insulin biosynthesis (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- TISSUE SPECIFICITY: Pancreatic beta cells.
 CC -!- SIMILARITY: Contains 1 homeobox domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; AF004431; AAB61665.1; -
 DR HSPSP; P13297; IIG7.
 DR TRANSFAC; T04297; -
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR009057; Homeodomain like.
 DR InterPro; IPR000047; HTH_lambrepreser.
 DR Pfam; PF000046; Homeobox; 1.
 DR PRINTS; PR00024; HOMEBOX.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX 1; 1.
 DR PROSITE; PS00071; HOMEBOX 2; 1.
 DR Developmental protein; DNA-binding; Homeobox; Nuclear protein.
 KW DOMAIN 49 61 Poly-Ser.

FT DOMAIN 119 132 Poly-Ser.
 FT DOMAIN 136 151 Poly-Ala.
 FT DOMAIN 169 174 Poly-Pro.
 FT DNA_BIND 237 296 Homeobox.
 FT DOMAIN 323 326 Poly-Asp.
 SQ SEQUENCE 365 AA; 37689 MW; C4AAB702D051F1F2 CRC64;
 Query Match 40.0%; Score 558.5; DB 1; Length 365;
 Best Local Similarity 41.5%; Pred. No. 8.4e-38;
 Matches 148; Conservative 26; Mismatches 80; Indels 103; Gaps 13;
 QY 1 MENLQGTFLNNTQLA---QFSEMKAPM-----COYSQVNSFYKLSPP--- 41
 DB 7 MEGFPGSAFLSSPPLAALHSMKMTPLYPAAYPPLPTGPPSSSSSSSSSSSPSPPLGA 66
 QY 42 -----GLGP-----OLAACTPHGTDILSRP-----VATPNS-- 69
 DB 67 HNPGLKPPAAGSLGSLGPPQOLSAATPHGINDILSRPMPVAGSALPSGSSSS 126
 QY 70 -----VGSFSGKAGNEYPTTRNCWADTGD--WRGSARPCGNTDPD---LSDTIHKKHTR 144
 DB 184 AVAAGVRYPKPLAELPGRTPIFWPGVQSPWDRALAC--TPHQGSILLDKDKKHKTR 241
 QY 145 PTFTGHOIFALEKTFEOTKYLAGPERARLAYSLGWTESQVQVWFQNRRTKWKKSALPS 204
 DB 242 PTFSGQQIFALEKTFEOTKYLAGPERARLAYSLGWTESQVQVWFQNRRTKWKKHAEMA 301
 QY 205 SSTPRAPGASGDRAASENE--DDYNKPLDPDSDEKIRILLKRAAFSVLSGA 259
 DB 302 TAKKKQDSETERLKGTSENEEDDDYNKPLDPNSDEKITQLLKKHSGGSLHLHA 358
 RESULT 6
 NK61_HUMAN
 ID NK61_HUMAN STANDARD; PRT; 367 AA.
 AC P78426;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Homeobox protein Nkx-6.1.
 GN Name=Nkx6-1; Synonyms=Nkx6a;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 SEQUENCE FROM N.A.
 RC TISSUE=Pancratic islets;
 RX MEDLINE=97237060; PubMed=9119408; DOI=10.1006/geno.1996.4569;
 RA Inoue H., Rudnick A., German M.S., Veille R., Donis-Keller H.,
 Permutt M.A.,
 RT "Isolation, characterization, and chromosomal mapping of the human
 Nkx6.1 gene (Nkx6A), a new pancreatic islet homeobox gene";
 RL Genomics 40:367-370 (1997).
 CC -!- FUNCTION: May be important for control of islet development and/or
 regulation of insulin biosynthesis.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- TISSUE SPECIFICITY: Pancreatic beta cells.
 CC -!- SIMILARITY: Contains 1 homeobox domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; U66799; AAD11962.1; -

01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 Homeodomain protein (Fragment).
 Name=NKX-6.1;
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus
 NCBI_TaxID=9031;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98202514; PubMed=9533954; DOI=10.1016/S0925-4773(98)00018-5;
 RA Qiu M., Shimamura K., Sussel L., Chen S., Rubenstein J.L.;
 "Control of anteroposterior and dorsoventral domains of Nkx-6.1 gene
 expression relative to other Nkx genes during vertebrate CNS
 development."; *Mech. Dev.* 72:77-88(1998).
 [2]
 RN SEQUENCE FROM N.A.
 RA Qiu M.S., Li G.Y., Rubenstein J.L.R.;
 Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 DR EMBL; AF102991; AAC83926.1; -;
 DR HSSP; P13297; 1IG7; -;
 DR TRANSFAC; T04270; -;
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR009057; Homeodomain like.
 DR Pfam; PF00046; Homeobox; 1.
 DR PRINTS; PR00024; HOMEBOX.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX 1; 1.
 DR PROSITE; PS00071; HOMEBOX 2; 1.
 KW DNA-binding; Homeobox; Nuclear protein.
 FT NON_TER 1
 SQ SEQUENCE 153 AA; 1723 MW; D7FA29C7CFAABBC7 CRC64;
 Query Match 30.8%; Score 430.5; DB 2; Length 153;
 Best Local Similarity 64.9%; Pred. No. 1e-27;
 Matches 87; Conservative 14; Mismatches 30; Indels 3; Gaps 2;
 QY 119 WRGSRPGC-NTPDPLSDTIHKKKTRPTFTTCHOIFALEKTEQTKYLAGPERARLAYSL 177
 DB 3 WRDARIGCAPHQGSILLKDGKGRKHTRPTFTSGQQIFALEKTEQTKYLAGPERARLAYSL 62
 QY 178 GMTESQVWVFNRRTKRKSALEPSSSTPRAPGAGSDRAASENE--DDEYNKPLDPD 235
 DB 63 GMTESQVWVFNRRTKRKSALEPSSSTPRAPGAGSDRAASENE--DDEYNKPLDPD 122
 QY 236 SDEKIRLLLRKH 249
 DB 123 SDEKIAQLKKHK 136
 RESULT 9
 Q96LR0 PRELIMINARY; PRT; 135 AA.
 AC Q96LR0;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein FLJ25169.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 RN SEQUENCE FROM N.A.
 RP

TISSUE=Brain;
 Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
 RA Fukuzumi Y., Fujimori Y., Komiyama M., Suzuki Y., Hata H.,
 RA Nakagawa K., Mizuno S., Morinaga M., Kawamura M., Sugiyama T.,
 RA Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A., Kawakami B.,
 RA Nagai K., Isogai T., Sugano S.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK057898; BAB71610.1; -;
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR009057; Homeodomain like.
 SQ SEQUENCE 135 AA; 14322 MW; 76B68B5AB913D6F CRC64;
 Query Match 27.3%; Score 381; DB 2; Length 135;
 Best Local Similarity 92.6%; Pred. No. 1.1e-23;
 Matches 75; Conservative 0; Mismatches 2; Indels 4; Gaps 1;
 QY 186 VWFONRRTKWKSALEPSSSTPRAPG-----GASGDRAASENEDDEYNKPLDPDSDDEKI 241
 DB 55 VWFONRRTKWKSALEPSSSTPRAPGAGAGAGAGGDRAPSENEDEYNKPLDPDSDDEKI 114
 QY 242 RLLLRKHRAAFSVLSLGAHSV 262
 DB 115 RLLLRKHRAAFSVLSLGAHSV 135
 RESULT 10
 Q8SXK1 PRELIMINARY; PRT; 480 AA.
 ID Q8SXK1;
 AC Q8SXK1;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE RE18506P.
 GN Name=HGTX;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunco J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 DR EMBL; AY089592; AAL90330.1; -;
 DR HSSP; P02833; 9ANT.
 DR Flybase; FBgn0040318; HGTX.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR009057; Homeodomain like.
 DR InterPro; IPR000047; HTH_lambrepresr.
 DR Pfam; PF00046; Homeobox; 1.
 DR PRINTS; PR00024; HOMEBOX.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX 1; 1.
 DR PROSITE; PS00071; HOMEBOX 2; 1.
 KW DNA-binding; Homeobox; Nuclear protein.
 SQ SEQUENCE 480 AA; 51093 MW; 1877E735CF36322C CRC64;
 Query Match 24.9%; Score 347; DB 2; Length 480;
 Best Local Similarity 42.6%; Pred. No. 3.4e-20;

65 HAAEMATAKXKQBELGDDGDCSE-----PMSDSE 95

Db

RESULT 13

Q9GLL7

ID Q9GLL7 PRELIMINARY; PRT; 80 AA.

AC Q9GLL7; (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)

DE Homeobox protein Nkx6.1 (fragment).

OS Sus scrofa (pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI_TaxID=9823;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Pancreas;

RA Wong R.Y., Tsch B.E.;

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).

DR EMBL; AF236157; AAG09467.1; --

DR HSP; P13297; IIG7.

DR GO; GO:0005634; C:nucleus; IEA.

DR GO; GO:0003700; F:transcription factor activity; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR InterPro; IPR001356; Homeobox.

DR InterPro; IPR009057; Homeobox-like.

DR InterPro; IPR000047; HTH_lambrepres.

DR Pfam; PF00046; Homeobox; 1.

DR PRINTS; PR00024; HOMEBOX.

DR PRINTS; PR00031; HTHREPRESS.

DR ProDom; PD000010; Homeobox; 1.

DR SMART; SM00389; HOX; 1.

DR PROSITE; PS00027; HOMEBOX_1; 1.

DR PROSITE; PS00071; HOMEBOX_2; 1.

KW DNA-binding; Homeobox; Nuclear protein.

FT NON_TER 1

FT TER 80

SQ SEQUENCE 80 AA; 9375 MW; 44DCAPD9ECAB9F23 CRC64;

Query Match 22.1%; Score 308; DB 2; Length 80;

Best Local Similarity 89.4%; Pred. No. 5.7e-18;

Matches 59; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 133 LSDTHKKKTRPTFTGHQIFALEKTEQTKYLAGPERARLAYSLGWTESQVKVWFQNR 192

Db 14 LLDKGRKKTRPTFTFSQQIFALEKTEQTKYLAGPERARLAYSLGWTESQVKVWFQNR 73

QY 193 TKWRKK 198

Db 74 TKWRKK 79

RESULT 14

Q81IK2

ID Q81IK2 PRELIMINARY; PRT; 106 AA.

AC Q81IK2;

DT 01-MAR-2003 (Tremblrel. 23, Created)

DT 01-MAR-2004 (Tremblrel. 23, Last sequence update)

DT 25-OCT-2004 (Tremblrel. 28, Last annotation update)

DE .COG-1B (Hypothetical protein R03C1.3b).

GN Name=cog-1; Synonyms=R03C1.3b; ORFNames=R03C1.3;

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=22370693; PubMed=12482710; DOI=10.1006/dbio.2002.0850;

RA Palmer R.E., Inoue T., Sherwood D.R., Jiang L.I., Sternberg P.W.;

RT "Caenorhabditis elegans cog-1 locus encodes GTX/Nkx6.1 homeodomain proteins and regulates multiple aspects of reproductive system

"Functional annotation of a full-length mouse cDNA collection";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Small intestine;
RA The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Small intestine;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Suganara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Small intestine;
RX MEDLINE=20530913; PubMed=11076961; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Small intestine;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito K., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Teijina Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK008173; BAC25203.1; -;
DR MGD; MGI:1352738; Nrx6-2.
SQ SEQUENCE 214 AA; 21930 MW; 7E10911A8BD42D52 CRC64;

Query Match 17.2%; Score 239.5; DB 2; Length 214;
Best Local Similarity 36.5%; Pred. No. 9.1e-12;
Matches 69; Conservative 20; Mismatches 77; Indels 23; Gaps 8;

QY 1 MESNLGGTLLNNTQLA---QFSEKAPMCQSVON--SFYKLSPPGLGQPLAAGPHGI 55
DB 1 MDANRFGAFVLSAPLAAHNNWEMKTSLPFPAALQGPAGKFTALGSLGAQLPLGTPHGI 60

QY 56 TDILSRFPVATPNSLLSGYPHVAGFGLSSQGVYGVQVQV---SFSKAGNEVPTTRNCW 112
DB 61 SDILGRFVGAAGGLLGLSLPRNLGLA--SSAGVYFGFAAARVAGYKPLAELPGRPIFW 118

QY 113 ADT--GQDWR-----GSARPCGNTPDPLSDTHKKHTRPTFTGHQIFALEKTFEOTKYL 165
DB 119 PGVVGSPWDRPLRAGSAQAGG-----VLDKDGKXKSRPTFGQDQLRAGEDFRADQVF 173

QY 166 APERARLA 174
DB 174 GRP-RAAA 181

RESULT 17

Q6E7C7 PRELIMINARY; PRT; 380 AA.
ID Q6E7C7
AC Q6E7C7; 2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE Distal-less.
OS Oikopleura dioica.
OC Eukaryota; Metazoa; Chordata; Urochordata; Appendicularia;
OC Oikopleuridae; Oikopleura.
OX NCBI_TaxID=34765;
RN [1]
RP SEQUENCE FROM N.A.
RA Sperber S., Seo H.-C., Mikhaileva J., Jensen M.F., Ekker M.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
CC 1- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR EMBL; AY532061; AAT39335.1; -;
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeobox_like.
DR InterPro; IPR000047; HTH_lambdarepressor.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00031; HTHREPRESSR.
DR ProDom; PD00010; Homeobox; 1.
DR SMART; SM00389; Hox; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 380 AA; 39387 MW; CFC615A2E9416F22 CRC64;

Query Match 17.1%; Score 239; DB 2; Length 380;

Best Local Similarity 32.9%; Pred. No. 2.1e-11;
Matches 70; Conservative 25; Mismatches 56; Indels 52; Gaps 9;

QY 43 LGPOLAAGTTHGTTDLSPVATPNSLLSGYPHVAGFGLSSQGVYGVQVGFSGKAGN 102
DB 2 LGSNETAGMSYGT-----GYPTA-YGAASRGSNYS--GYNNSSS 41

QY 103 EYP-----TRTRNCWADTGQWRGSA-RPCGNTPDPLSDTHKKHTRP- 145
DB 42 SYPGSSLLNSPTGGLAASQYAAFAOSGYPPLGFADNPFGLLPDSKG---KKKMKRKP 98

QY 146 -TFTGHQIFALEKTFEOTKYLGPERRARLAYSLGNTESQVKVWFQNRRTKWRK----- 197
DB 99 TIYSLQLQELNKPNTQYALPERELAASLGLTQVQKWFQNRRSKYKIMKQGV 158

QY 198 KSALEPSSSTPRAPGASGDRAASENEDDEYNK 230
DB 159 SVKIENDGGSP-----AGSGDEAGANGEDDEENQ 187

RESULT 18

HMHP CHICK
ID HMHP CHICK STANDARD; PRT; 277 AA.
AC Q05502;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Homeobox protein PRH.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9308175; PubMed=1360645;
RA Crompton M.R., Bartlett T.J., Macgregor A.D., Manfioletti G.,
RA Buratti E., Giannotti V., Goodwin G.H.;

RT "Identification of a novel vertebrate homeobox gene expressed in
 RT hematopoietic cells.";
 RL Nucleic Acids Res. 20:5661-5667(1992).
 CC -!- FUNCTION: Recognizes the DNA sequence 5'-ATTAA-3'. May play a role
 CC in hematopoietic differentiation.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- TISSUE SPECIFICITY: In all hematopoietic tissues except peripheral
 CC blood erythrocytes and in the liver and lung.
 CC -!- SIMILARITY: Contains 1 homeobox domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X64711; CAA45966.1; --
 CC FIR; S78063; S78063.
 CC TRANSFAC; T02091; --
 CC InterPro; IPR001356; Homeobox.
 CC Pfam; PF00046; Homeobox; 1.
 CC PRINTS; PR00024; HOMEBOX.
 CC ProDom; PD000010; Homeobox; 1.
 CC SMART; SM00389; HOX; 1.
 CC PROSITE; PS00027; HOMEBOX_1; 1.
 CC PROSITE; PS50071; HOMEBOX_2; 1.
 CC Developmental protein; DNA-binding; Homeobox; Nuclear protein.
 KW DOMAIN 1 140 Pro-rich.
 FT DNA BIND 144 203 Homeobox.
 SQ SEQUENCE 277 AA; 30213 MW; BE744C143FB9F9FC CRC64;
 Query Match 16.3%; Score 228; DB 1; Length 277;
 Best Local Similarity 30.0%; Pred. No. 1.1e-10; Indels 72; Gaps 11;
 Matches 79; Conservative 25; Mismatches 87;
 QY 25 PNCQSVQNSVFYKLPGLPQQAAGTGHGIDILSRVPATPNS---SLLSGY----- 74
 DB 25 PLLQPAHPFPFYIEDILGRG-AAAPAPSLPAPPPPTLPSNSFTSLVAFYRPPVPEP 83
 QY 75 -----PHVAGFGLSSQGVYGVGVGSPKAGNEY-----PFRTRNCWADTQGD 118
 DB 84 TPIHAFSHLAATYG---TCGAYAGP-LYSFPRAVDGYTHALIRQDPLGKPLLSFFIQ- 138
 QY 119 WRGSRPCGNTDPLSDTIHKKHTRPTTGHQIFALEKTPFTQTKYLAGPEARLAYSIG 178
 DB 139 -----RP-----LHKRGQGVRFNSQTELEKKFTQKYLSPPRKRRLAKULQ 182
 QY 179 MTESQVKVWFQNRRTKWRKSALEP-----SSSTPRAPGASGD--- 217
 DB 183 LSERQVKTQFQNRARWRKRLQENQATKKEAEGTGDHGDPRSEGSFSPAGGEARQD 242
 QY 218 -RAASENEDDEYNKPLDPDSDE 239
 DB 243 SPSAASQEDPE--SDVSDSDSQE 263
 RESULT 19
 TLX2 HUMAN
 ID TLX2 HUMAN STANDARD; PRT; 284 AA.
 AC O43763; Q9UQ48;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE T-cell leukemia homeobox protein 2 (Homeobox protein Hox-11L1) (Neural
 DE crest homeobox protein).
 GN Name=TLX2; Synonyms=HOX11L1, NCX;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Delgado P., Rodriguez R.E., Gonzalez-Sarmiento R.;
 RT "Genomic characterization and chromosomal location of the human
 RL homeobox gene HOX11L1.";
 RN Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99377083; PubMed=10446220; DOI=10.1074/jbc.274.34.24401;
 RA Itekhia Y., Shimizu H., Kang M.M., Sasagawa K., Sekiya S.,
 RA Tokuhisa T., Hatano M.;
 RT "An enhancer element for expression of the Ncx (Enx, Hox11L1) gene in
 RT neural crest-derived cells.";
 RL J. Biol. Chem. 274:24401-24407(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Frange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: Contains 1 homeobox domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ002607; CAA05636.1; --
 CC EMBL; AJ002608; CAA05636.1; JOINED.
 CC EMBL; AJ002609; CAA05636.1; JOINED.
 CC EMBL; AB008501; BAA83463.1; --
 CC EMBL; BC006357; BAA06356.1; --
 CC HSP; P13297; IIG7.
 CC TRANSFAC; T04367; --
 CC Genew; HGNC:5057; TLX2.
 CC MIM; 604240; --
 CC InterPro; IPR001356; Homeobox.
 CC InterPro; IPR009057; Homeodomain_like.
 CC Pfam; PF00046; Homeobox; 1.
 CC PRINTS; PR00024; HOMEBOX.
 CC ProDom; PD000010; Homeobox; 1.
 CC SMART; SM00389; HOX; 1.
 CC PROSITE; PS00027; HOMEBOX_1; 1.
 CC PROSITE; PS50071; HOMEBOX_2; 1.
 CC Developmental protein; DNA-binding; Homeobox; Nuclear protein.
 KW DOMAIN 27 115 Gly-rich.
 FT DNA BIND 157 216 Homeobox.
 FT CONFLICT 16 16 P -> A (in Ref. 1).
 FT CONFLICT 30 32 TPG -> PR (in Ref. 1).
 FT CONFLICT 37 48 LGRGQGHGENG -> WYAGQVIGWEA (in Ref. 1).

[illegible]

DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Homeobox protein NK7.1.
GN Name=NK7.1;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon R;
RA Sakoyama Y., Miata I., Smith R.P.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC EMBL; AF054614; AAC08423.1; -;
DR HSP; P14653; 1B72.
DR FlyBase; FBgn0024321; NK7.1.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; P:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 641 AA; 70310 MW; A6DB34E0FC6F3DA CRC64;
Query Match 15.8%; Score 220; DB 2; Length 641;
Best Local Similarity 32.9%; Pred. No. 1.5e-09;
Matches 70; Conservative 20; Mismatches 63; Indels 60; Gaps 8;
QY 37 KLS-----PPGLGQLAAGTGHITDILSRPVATPNSLLSGYPHVGGLSSGGVYG 91
DB 327 KLSSTVALPFDI-----SPGTGSDSLMRDKLMANNS-----SSPGSNVN 365
QY 92 PQVGSFSKAGNEYPTRNCWADTGQWRGSRPCGNTPDPLSDTIHKKHTPTFTGHO 151
DB 366 AQMS-----NANSTLETTED-----DSDSGSTARRKKKARTFTGRQ 404
QY 152 IFALEKTFQTKYLAGPERARLAYSIGMTESQVWFQNRRTKWRK-----SALEPSS 205
DB 405 IFELERWFENKKYLSASERTENAKLLMVTETQVKLWFQNRRTKWKQKQDNVTNNEAEHKS 464
QY 206 STPRAPGASGDRAAENEDDEVNKLDPDSD 238
DB 465 S--NAKPGATGTATTTPS-----GEPTDKRSSN 490
RESULT 25
Q9VFK4 PRELIMINARY; PRT; 721 AA.
AC Q9VFK4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE CG8524-PA (CG8524-pb).
GN Name=NK7.1; ORFNames=CG8524;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20196006; PubMed=107311132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.E., Yandell M.D., Zhang Q., Chen L.X.,

Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Aril J.P., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Balaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fiesler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Heston N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA LaRoche P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]


```

RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR EMBL; AE003700; AAF55051.2; -.
DR HSSP; PI4653; 1B72.
DR IntAct; Q9VFK4; -.
DR FlyBase; FBgn0024321; NK7.1.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; P:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeodomain like.
DR InterPro; IPR011051; Rmc1 like_cupin.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEOBOX.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEODOMAIN_1; 1.
DR PROSITE; PS00071; HOMEODOMAIN_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 721 AA; 78915 MW; 25A3C5785C3BDBA5 CRC64;

Query Match 15.8%; Score 220; DB 2; Length 721;
Best Local Similarity 32.9%; Pred. No. 1.8e-09;
Matches 70; Conservative 20; Mismatches 63; Indels 60; Gaps 8;

QY 37 KLS-----PPGLGPOLAACTPHGTTDILSRPVATPNSLLSGYPHVGFGSLSSQGVYVG 91
DB 327 KLSSTVALPPDIT-----SPTGSSDLSMRDXLMANNS-----SSPGSNVN 365
QY 92 POGVSFSKAGNEYPTRTRNCWADTQGDWRGSRPCGNTPDPLSDITHKKCHTRPTFTGHO 151
DB 366 AQMQS-----NANSTLETED-----DSDSGSTDARRKKKARTFTGRQ 404
QY 152 IPALEKTPQOTKYLAGPERALAYSIGMTESQVKVFWQNRRTKWKK-----SALEPSS 205
DB 405 IPELEKMFENKYLASERTEAKLLMVTQVKIWFQNRRTKWKKQDVTNNEAAEHKS 464
QY 206 STPRAPGAGSDRAASENEDDYNKPLDPDSD 238
DB 465 S--NAKPGATGATTTPTS-----GPTDKRSSN 490

RESULT 26
QYKVF4 PRELIMINARY; PRT; 322 AA.
AC Q7KVF4;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE CG3629-PB.
GN Name=Dil; ORFNames=CG3629;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
ON [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadiieu E., Center A., Chandra I.,
RA Chertiy J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

```

```

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskaas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RL melanogaster euchromatinic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskaas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RL a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RL systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR EMBL; AE003465; AAF47280.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; P:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeodomain like.
DR InterPro; IPR011051; Rmc1 like_cupin.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEOBOX.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEODOMAIN_1; 1.
DR PROSITE; PS00071; HOMEODOMAIN_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 721 AA; 78915 MW; 25A3C5785C3BDBA5 CRC64;

```



```

DR InterPro; IPR009057; Homeobox domain like.
DR InterPro; IPR000047; HTH lambrapressr.
DR Pfam; PF00046; Homeobox_1.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00031; HTHREPRESSR.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
DR DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 322 AA; 34710 MW; C9E2F9FASC7FFDE CRC64;

Query Match 15.7%; Score 219.5; DB 2; Length 322;
Best Local Similarity 34.6%; Pred. No. 6.9e-10;
Matches 55; Conservative 29; Mismatches 50; Indels 25; Gaps 5;

QY 78 AGFGLSSQGVYGPQVGS-----FSKAGNEYPTTRNCWADTGDWRGSRPCGNTP 130
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 43 AGYGIRSTYQHFGPQGQDGFPSRSGALGYPPPMHQNYS--SGYHLGSYAPPCASPP 100
QY 131 D---PLSDTIH-----KKHTRPTFTGHOIFALEKTFEOTKYLAGEPARLAYS 176
DB 101 KDFESISDKCEDSLRVNGKKGKMKRPTIYSSLOQLNRRFORTQYLALPERALAA 160
QY 177 LGMTESQVKWFQNRRTKWK--KSALEPSSSTPRAPGG 213
DB 161 LGLTQTQVKIWFQNRRSKYKQWKAAGQGTSNGLPGG 199

RESULT 28
QY Q9W028 PRELIMINARY; PRT; 327 AA.
ID AC Q9W028; Q9W029;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CG3629-PA
GN Name=Dll; ORFNames=CG3629;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gockayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abell J.F., Agbayani A., An H.J., Andrews-Franckoch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Bensch P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,

```


DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOK; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 333 AA, 34396 MW, 15.6% Score 218; DB 2; Length 333;
Best Local Similarity 41.4%; Pred. No. 9.5e-10;
Matches 48; Conservative 21; Mismatches 39; Indels 8; Gaps 3;
Query Match 15.6%; Score 218; DB 2; Length 333;
Best Local Similarity 29.9%; Pred. No. 9.5e-10;
Matches 73; Conservative 29; Mismatches 78; Indels 64; Gaps 9;
QY 120 RGSARPCGNTPDPLSDTHKKHTRPTTGHQIFALEKTPQTKYLAGPERARLAYSLGM 179
DB 21 RGRA-----DSEKFTNALERKKKRTTFSGQIFELERQFKQKYLSSGERAEVANQLNV 76
QY 180 TESQVKVWFQNRRTKWRKK---SALEPSSSTPRAPGGAS-CDRAASENEDDEYNKP 231
DB 77 TETQVKVWFQNRRTKWKQENISNTEAQSIQOTESQAKIAETVTLKREDSETSQP 132
RESULT 30
Q9OX99 PRELIMINARY; PRT; 333 AA.
AC Q9OX99;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Transcription factor (T-cell leukemia, homeobox 1).
GN Name=TLX1; Synonyms=Hox11;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N;
RA Koehler K., Dear N.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Salivary gland;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Krzywinski M.I., Skalka U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Salivary gland;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR EMBL; AJ251787; CAB63267.1; -;
DR EMBL; BC018246; AAH18246.1; -;
DR HSP; P13297; IIG7.
DR MGD; MGI:98769; Tlx1.
DR GO; GO:0007417; P:Central nervous system development; IMP.
DR GO; GO:0007515; P:Lymph gland development; TAS.
DR GO; GO:0030182; P:neuron differentiation; IMP.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR003057; Homeobox.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.

DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOK; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 333 AA, 34396 MW, 15.6% Score 218; DB 2; Length 333;
Best Local Similarity 29.9%; Pred. No. 9.5e-10;
Matches 73; Conservative 29; Mismatches 78; Indels 64; Gaps 9;
QY 36 YKLSPPGLGQLAAGTHGITDILSRPVATPNSLLSGYPHVAGFG-----LSSQ 86
DB 74 YGAGGPG-GPGGAGGAGGACSMGLPFGSYNNMALAGGPGGCGGGGAGAGALSA 132
QY 87 GVV-----YGPQVGSFSK-----AGNEYPTTRNCWADTCQDW 119
DB 133 GVTRVPAHRPLAGAVAHQPQLTGVPSVPAVGVNLTGLTFP-----WMSNRY 186
QY 120 ---RGSARPCGNTPDPLSDTHKKHTRPTTGHQIFALEKTPQTKYLAGPERARLAY 176
DB 187 TKDRFTGHPYQNRTPP-----KCKKPRTSFTRLQICELEKRFHROKYLASAERAA 240
QY 177 LGMTESQVKVWFQNRRTKWRKKSALEPSSSTPRAPGGASCDRAASENEDDEYNK----- 230
DB 241 LKMTDAQVKTWFQNRRTKWRRTAEERAEQQA-----NRILLQQQEAFOKSLAQL 294
QY 231 PLDP 234
DB 295 PADP 298
RESULT 31
HMPH MOUSE STANDARD; PRT; 271 AA.
ID HMPH MOUSE
AC P43120;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Homeobox protein PRH (Hematopoietically expressed homeobox) (Homeobox protein HEX).
DB Name=Hex; Synonyms=Prhx;
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=93219086; PubMed=8096636;
RA Bedford F.K., Ashworth A., Enver T., Wiedemann L.M.;
RT "HEX: a novel homeobox gene expressed during haematopoiesis and conserved between mouse and human."
Nucleic Acids Res. 21:1245-1249(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Liver;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Krzywinski M.I., Skalka U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Salivary gland;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR EMBL; AJ251787; CAB63267.1; -;
DR EMBL; BC018246; AAH18246.1; -;
DR HSP; P13297; IIG7.
DR MGD; MGI:98769; Tlx1.
DR GO; GO:0007417; P:Central nervous system development; IMP.
DR GO; GO:0007515; P:Lymph gland development; TAS.
DR GO; GO:0030182; P:neuron differentiation; IMP.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR003057; Homeobox.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Recognizes the DNA sequence 5'-ATTAA-3'. May play a role
CC in hematopoietic differentiation.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- DEVELOPMENTAL STAGE: Expressed during hematopoiesis.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z21524; CAA79729.1; -.
CC EMBL; BC057986; AAH57986.1; -.
CC PIR; S30230; S30230.
CC TRANSFAC; T03417; -.
CC MGD; MGI:96086; Hhex.
CC GO; GO:0005634; C:nucleus; IDA.
CC GO; GO:0007420; P:brain development; IMP.
CC InterPro; IPR001356; Homeobox.
CC InterPro; IPR009057; Homeodomain_like.
CC Pfam; PF00046; Homeobox; 1.
CC PRINTS; PR00024; HOMEBOX.
CC Prodom; PD00010; Homeobox; 1.
CC PROSITE; PS00027; HOMEBOX 1; 1.
CC PROSITE; PS00071; HOMEBOX 2; 1.
KW Developmental protein; DNA-binding; Homeobox; Nuclear protein.
FT DOMAIN 1 134 Pro-rich.
FT DNA_BIND 138 197 Homeobox.
FT SEQUENCE 271 AA; 29986 MW; 376765557A71C962 CRC64;

Query Match 15.6%; Score 217.5; DB 1; Length 271;
Best Local Similarity 30.9%; Pred. No. 8.1e-10;
Matches 82; Conservative 19; Mismatches 65; Indels 99; Gaps 12;

QY 46 QLAAGTGHGTTILSR-----PVATN-----SLLSGY-----PH 76
Db 26 OPAHTPFYIDILGRGAAPTPTTLPSPNSFTSLVSSYRTVPVPTVHPAFSHHPA 85
QY 77 VA-----GGGLSSQGVYGVQVGSFKAGNEY-----PRTTNCWADTGQDWR 120
Db 86 ALAAAYGSGGFG-----PLYPPFTVNDYTHALLRHDPLGKPLLSPPFLQ--- 132
QY 121 GSARPCGNTPDPLSDTIHKKHTRPTFTGHQIFALEKTFEQT KYLAGPERARLAYSLGWT 180
Db 133 ---RP-----LHKRKGQVRFNSDQIVLEKKFETQKYLSPERKELAKWLQLS 178
QY 181 ESQVKWFQNRKTR-----KXSALEPSSSTPRAPG-----GASGDR--- 218
Db 179 ERQVTFWFQNRKARLQENPQSNKKDALD-SLDTSCQQLDLPFSEQNGKASLDLSQC 237
QY 219 ----AASNEDEYNKPLDPSDDSE 239
Db 238 SPSPASQEDPDSSEDSQEVDIIE 262

RESULT 32
Q9CRV1 PRELIMINARY; PRT; 280 AA.
AC Q9CRV1
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 13 days embryo head cDNA, RIKEN full-length enriched
DE library, clone:310032E02 product:HOMEBOX PROTEIN PRH (HOMEBOX)

DE DE
GN Name=Hhex;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RL "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RL "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RA The FANTOM Consortium.
RL the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaki S., Inoue K., Togawa Y., Izawa E., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC EMBL; AK014111; BAB29163.1; -.
CC HSP; P22808; INK3.
CC MGD; MGI:96086; Hhex.
CC GO; GO:0005634; C:nucleus; IDA.
CC GO; GO:0007420; P:brain development; IMP.
CC GO; GO:0009887; P:organogenesis; IMP.
CC

DR	GO:0042127; P:regulation of cell proliferation; IMP.	DR	ProDom; PD000010; Homeobox; 1.
DR	InterPro; IPR001356; Homeobox.	DR	SMART; SM00389; HOX; 1.
DR	InterPro; IPR009057; Homeodomain_like.	DR	PROSITE; PS00027; HOMEBOX 1; 1.
DR	Pfam; PF00046; Homeobox; 1.	DR	PROSITE; PS00071; HOMEBOX_2; 1.
DR	PRINTS; PR00024; HOMEBOX.	KW	DNA-binding; Homeobox; Nuclear protein.
DR	ProDom; PD000010; Homeobox; 1.	QV	SEQUENCE 292 AA; 32486 MW; 11075B21680ACD79 CRC64;
DR	SMART; SM00389; HOX; 1.		
DR	PROSITE; PS00027; HOMEBOX 1; 1.		
DR	PROSITE; PS00071; HOMEBOX 2; 1.		
KW	DNA-binding; Homeobox; Nuclear protein.		
FT	NON_TER 1		
SQ	SEQUENCE 280 AA; 30771 MW; 9405DC67E1842FF0 CRC64;		
QV	Query Match 15.6%; Score 217.5; DB 2; Length 280;	QV	Query Match 15.6%; Score 217.5; DB 2; Length 292;
Db	Best Local Similarity 30.9%; Pred. No. 8.4e-10;	Db	Best Local Similarity 29.0%; Pred. No. 8.9e-10;
	Matches 82; Conservative 19; Mismatches 65; Indels 99; Gaps 12;		Matches 75; Conservative 36; Mismatches 91; Indels 57; Gaps 11;
QV	46 QLAACHTPHGIDILSR-----PVATN-----SSLLSGV-----PH 76	QV	2 ESNLQCTFLNNTQLAQF--SEMKAPMCQVSVQNSFYKLSPPGLGQQL-----AAG 50
Db	35 QPAHPTFFVIDILGRGAAPTPTLPSPNSSFTLSVSVTPVYETPVHPAFSHHPA 94	Db	33 ENSQSASVTRGSESSYLGSPISRPSPAPYCLSANF---PAGIGATFDESGYSVNVSL 89
QV	77 VA-----GFGLSGQGVYGPQVGSFKAGNEY-----PTRTRNCWADTGQDWR 120	QV	51 TPHGITDI-LSRVVAPNSLLSGYPHVAGFGLSSQGVYGPQVGS-----FSKAG 101
Db	95 ALAAAYGSGRGG-----FLYPFRTVNDYTHALLRHDPGLGKPLLWSPFLQ--- 141	Db	90 APAGVTRVPARHPIPGAPISAFIPDAVPGGLGSLNF-----PWIESRRFVKERTAAA 143
QV	121 GSARPCGNTPPDLSTIHKXKTRFTTGHQIFALEKTFEQTKYLAGERARLAYSLGWT 180	QV	102 NEYP-TETRNCWADTGQDWRGSRAPCGNTDPDLSTIHKXKTRFTTGHQIFALEKTFE 160
Db	142 ---RP-----LHKRKGQVRFNDQTVLEKKFETQKYLSPPERKRLAKMLQLS 187	Db	144 ALTPFTVTRI-----GHPYQNRTP-----KTKKPTSFVRVQICELEKRFH 186
QV	181 ESQVKVWFQNRRTKWR-----KKSALPSSSTPRAPG-----CASGDR--- 218	QV	221 SENEDDEYNKPL---DPD 235
Db	188 ERQKTVFQNRRAKWRRLKQENPQSNKKDALD-SLDTSCQGLPSEQNKGASLDRSQC 246	Db	241 IQLQHDFAFQKLSDSIQPD 259
QV	219 -----AASENEDDEYNKPLDPDSDE 239		
Db	247 SPSPASQEDPDSEISRDSQEVNDIE 271		
RESULT 33		RESULT 34	
Q9DFC2		TLX3 CHICK	
ID	Q9DFC2 PRELIMINARY; PRT; 292 AA.	ID	TLX3 CHICK STANDARD; PRT; 297 AA.
AC	Q9DFC2	AC	O93367;
DT	01-MAR-2001 (TrEMBLrel. 16, Created)	DT	15-JUL-1999 (Rel. 38, Created)
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)	DT	15-JUL-1999 (Rel. 38, Last sequence update)
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)	DT	05-JUL-2004 (Rel. 44, Last annotation update)
DE	Homeobox protein Hox11L2.	DE	T-cell leukemia homeobox protein 3 (Homeobox protein Hox-11L2)
OS	Xenopus laevis (African clawed frog).	DE	(Homeobox TLX-3).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	GN	Name=TLX3; Synonyms=HOX11L2;
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;	OS	Gallus gallus (Chicken).
OC	Xenopodinae; Xenopus.	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	NCBI_TaxID=8355;	OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
RN	SEQUENCE FROM N.A.	OC	Gallus.
RP	Patterson K.D., Krieg P.A.;	OX	NCBI_TaxID=9031;
RX	DOI=10.1002/(SICI)1097-0177(199901)214:1<34::AID-DVNDY4>3.3.CO;2-I;	RP	SEQUENCE FROM N.A.
RA	Patterson K.D., Krieg P.A.;	RC	STRAIN=Rhode Island red;
RT	"Hox11-family genes Hox11 and Hox11L2 in xenopus: Hox11L2	RC	MEDLINE=98315212; PubMed=9651221;
RT	expression is restricted to a subset of the primary sensory neurons."	RA	Logan C.C., Wingate R.J.T., McKay I.J., Lumsden A.;
RL	Dev. Dyn. 214:34-43(1999).	RT	"Tlx-1 and Tlx-3 homeobox gene expression in cranial sensory ganglia
RL	[2]	RT	and hindbrain of the chick embryo: markers of patterned
RP	SEQUENCE FROM N.A.	RL	connectivity";
RA	Patterson K.D., Krieg P.A.;	RL	J. Neurosci. 18:5389-5402(1998).
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.	CC	-!- FUNCTION: Seems to be involved in the development of cranial
CC	-!- SUBCELLULAR LOCATION: Nuclear (By similarity).	CC	sensory innervation from peripheral ganglia.
CC	EMBL; AF283693; AAG1452.1; --	CC	-!- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR	HSSP; P13297; 1IG7.	CC	-!- TISSUE SPECIFICITY: Expression is restricted to neurones in the
DR	GO:0005634; C:nucleus; IEA.	CC	peripheral and central nervous system.
DR	GO:0003700; F:transcription factor activity; IEA.	CC	-!- SIMILARITY: Contains 1 homeobox domain.
DR	GO:0006355; P:regulation of transcription, DNA-dependent; IEA.	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
DR	InterPro; IPR001356; Homeobox.	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
DR	InterPro; IPR009057; Homeodomain_like.	CC	the European Bioinformatics Institute. There are no restrictions on its
DR	Pfam; PF00046; Homeobox; 1.	CC	use by non-profit institutions as long as its content is in no way
DR	PRINTS; PR00024; HOMEBOX.	CC	modified and this statement is not removed. Usage by and for commercial
DR	ProDom; PD000010; Homeobox; 1.	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
DR	SMART; SM00389; HOX; 1.	CC	or send an email to license@isb-sib.ch).
DR	PROSITE; PS00027; HOMEBOX 1; 1.	CC	-----
DR	PROSITE; PS00071; HOMEBOX 2; 1.	CC	EMBL; AF071875; AAC23901.1; --
KW	DNA-binding; Homeobox; Nuclear protein.	CC	HSSP; P13297; 1IG7.
FT	NON_TER 1		
SQ	SEQUENCE 280 AA; 30771 MW; 9405DC67E1842FF0 CRC64;		

Thu Apr 14 08:26:22 2005

```
DR InterPro; IPR001827; Antennapedia.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeodomain_like.
DR InterPro; IPR000047; HTH_lambrepreser.
DR Pfam; PF00046; Homeobox_1.
DR PRINTS; PR00025; ANTENAPEDIA.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00031; HTHREPRESSR.
DR ProDom; PD00010; Homeobox; 1.
DR PROSITE; PS00032; ANTENAPEDIA; FALSE_NEG.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Activator; Developmental protein; Diabetes mellitus; DNA-binding;
Homeobox; Nuclear protein; Phosphorylation; Transcription regulation.
FT DOMAIN 13 73 Transactivation domain (by similarity).
FT SITE 42 48 Poly-Pro.
FT SITE 119 124 Antip-type hexapeptide.
FT DOMAIN 198 204 Nuclear localization signal.
FT DNA_BIND 147 206 Homeobox.
FT DOMAIN 217 221 Poly-Gly.
SQ SEQUENCE 284 AA; 30999 MW; A57D04569D14E3C4 CRC64;

Query Match 15.5%; Score 217; DB 1; Length 284;
Best Local Similarity 30.3%; Pred. No. 9.4e-10;
Matches 70; Conservative 26; Mismatches 95; Indels 40; Gaps 9;

QY 25 PNCQSVNSFKLSPGLGQ-----LAAGTGHGTDILSRPVATPNSLLSGYPHV 77
DB 33 PACLY-----MGRQPPPPPPQFTSSLSGLSQSPDIPVEVPLASDDPAGHLHHLL 87
QY 78 AGFGGLSSQGVYGVQVGSFSGKAGNEYTRTR-----NCWADTGDWRGSRPC 126
DB 88 PAQLGLAHPPP--GFPNGTEPGLEENRVLPFPNWKSTKAHAW--KGQ-WAGGA--- 139
QY 127 GNTDPDLSDTIHKKHTRPTFTCHQIFALEKTFEOTKYLAGEPARARLAYSIGMTESQVKV 186
DB 140 -YTAEP-----EENKTRTAVTRAQLLEKEFLFNKYISRRVELAVMLNLTERHIKI 193
QY 187 WFNQRRTKWRKSKALEPSSSTPRAPGGA---SGDRAAENEDEYNKPLDP 234
DB 194 WFNQRRMKWKBEDKKRSGTSPGGGGEPEQDCAVTSGEELLAVPLPP 244

RESULT 36
Q9RLX2 PRELIMINARY; PRT; 271 AA.
AC Q9RLX2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hex (Prt) protein.
GN Names=Hex; Synonyms=Hex (Prt);
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129; TISSUE=Liver;
RA Myint Z., Inazu T., Tanaka T., Yamada K., Keng V.W., Inoue Y.,
RA Kuriyama M., Noguchi T.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR EMBL; AB017132; BAA76714.1;
DR EMBL; AB017130; BAA76714.1; JOINED.
DR HSSP; P22808; INK3.
DR MGD; MGI:96086; Hhex.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0007420; P:brain development; IMP.
DR GO; GO:0009887; P:organogenesis; IMP.
DR GO; GO:0042127; P:regulation of cell proliferation; IMP.
DR InterPro; IPR001356; Homeobox.
```

```
DR InterPro; IPR009057; Homeodomain_like.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD00010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 271 AA; 29952 MW; 4C0AD438CCAAAGF4 CRC64;

Query Match 15.5%; Score 216; DB 2; Length 271;
Best Local Similarity 30.9%; Pred. No. 1.1e-09;
Matches 82; Conservative 19; Mismatches 65; Indels 99; Gaps 12;

QY 46 QLAAGTGHGTDILSR-----PVATPN-----SSLLSGY-----PH 76
DB 26 QPAHPTPLYDDILGRGPAAPTFTPLPSNSSFTSLVSSYRTFVVEPTVHPAFSHHPA 85
QY 77 VA-----GFGGLSSQGVYGVQVGSFSGKAGNEY-----PTRTRNCWADTGDWR 120
DB 86 AALAAAYGPSGFGG-----PLYPPRTVNDYTHALLRHDPLGKPLLLSPFLQ--- 132
QY 121 GSARPCGNTDPLSDTIHKKHTRPTFTCHQIFALEKTFEOTKYLAGEPARARLAYSIGMT 180
DB 133 ---RP-----LHKRGQVRFNSNDQTVELEKFKFTQKYLSPERKRLAKMLQLS 178
QY 181 ESQVKVWFQNRRTKWR-----KSALEPSSSTPRAPG-----GASGDR--- 218
DB 179 ERQVKTWFQNRRAKWRLLKQENPQSNKKDALD-SLDTSCGQDLPSEQNKASLDRSQ 237
QY 219 ----AASENDEDEYNKPLDPDSDE 239
DB 238 SPSPASQEDPDSEISDSQDEVDIE 262

RESULT 37
TLX1 MOUSE
ID TLX1 MOUSE STANDARD; PRT; 332 AA.
AC P43345;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE T-cell leukemia homeobox protein 1 (Homeobox protein Hox-11) (Homeobox
DE TLX-1).
GN Name=TLX1; Synonyms=Hox11, Tlx-1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94206842; PubMed=7908826; DOI=10.1016/0925-4773(93)90016-Q;
RA Raju K., Tang S., Dube I.D., Kamel-Reid S., Bryce D.M., Breitman M.L.;
RT "Characterization and developmental expression of Tlx-1, the murine
RT homolog of Hox11.";
RL Mech. Dev. 44:51-64 (1993).
RN [2]
RP FUNCTION.
RX MEDLINE=94203286; PubMed=7908720; DOI=10.1038/368747a0;
RA Roberts C.W.M., Shutter J.R., Korsmeyer S.J.;
RT "Hox11 controls the genesis of the spleen.";
RL Nature 368:747-749 (1994).
RN [3]
RP DNA-BINDING.
RX MEDLINE=95319935; PubMed=7596820;
RA Tang S., Breitman M.L.;
RT "The optimal binding sequence of the Hox11 protein contains a
RT predicted recognition core motif.";
RL Nucleic Acids Res. 23:1928-1935 (1995).
CC -!- FUNCTION: Controls the genesis of the spleen. Binds to the DNA
CC sequence 5'-GGCGTAAAGTGG-3'.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: Expressed in various embryonic tissues,
```


CC including branchial arches, some component of the nervous system
CC and spleen.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S70632; AAB30542.1; -
CC EMBL; S70756; AAB30542.1; JOINED.
CC EMBL; S70629; AAB30542.1; JOINED.
CC HSP; P13297; IIG7.
CC TRANSFAC; T02055; -
CC MGD; MGI:98769; Tlx1.
CC GO; GO:0007417; P:central nervous system development; IMP.
CC GO; GO:0030182; P:neuron differentiation; IMP.
CC InterPro; IPR001356; Homeobox.
CC Pfam; PF00046; Homeobox; 1.
CC PRINTS; PR00024; HOMEBOX.
CC ProDom; PD000010; Homeobox; 1.
CC SMART; SM00389; HOX; 1.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS50071; HOMEBOX_2; 1.
CC Developmental protein; DNA-binding; Homeobox; Nuclear protein.
FT DNA BIND 203 262 Homeobox.
SQ SEQUENCE 332 AA; 34644 MW; 79FC71E2E44E6F82 CRC64;

Query Match 15.4%; Score 215.5; DB 1; Length 332;
Best Local Similarity 30.0%; Pred. No. 1.5e-09;
Matches 73; Conservative 30; Mismatches 77; Indels 63; Gaps 9;

QY 36 YKLSPPGLGQPLAAGTPHGITDILSRVATPNSSLLSGYPHVGFGG-----LSSQG 87
DB 74 YGAGGFG-GPGGAGGGGACSMGLPGSYNNMDLAGGPGGPGGGGAARRALSAG 132

QY 88 VY-----YGPQVGSFK-----AGNEYPTTRNCWADTGDW- 119
DB 133 VIRVPAHRPLAGAVHPQPLATGLTPVSPVAPGVNLTGLTP-----WMENRYT 186

QY 120 -RGSARCGNTDPPLSTIHKKHTRPTTGHQIFALEKTEQTKYLAGEPARLAYSL 177
DB 187 KDRFTGLPYQNRTP-----KKKPRTSFTRLQICELEKPRHQKYLASAERAAALAKAL 240

QY 178 GMTESQVQVWFQNRRTKWRKKSALFSSSTPPAPCGASGDRAASENDEYKN-----P 231
DB 241 KMTDAQVKTWFQNRRTKWRRTAEREAESEA-----NRLLQLQEEAFQKSLAQPLP 294

QY 232 LDP 234
DB 295 ADP 297

RESULT 38
Q9W710
ID Q9W710 PRELIMINARY; PRT; 388 AA.
AC Q9W710; 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Hlx-1 transcription factor.
GN Name=Hlx-1;
OS Pleurodeles waltlii (Iberian ribbed newt).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidae; Salamandridae;
OC Pleurodeles.
OX NCBI_TaxID=8319;
RN [1]
RP SEQUENCE FROM N.A.

RA Nicolas S., Le Parco Y.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR EMBL; AF106694; AAD2047.1; -.
DR HSP; P02836; IENH.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeobox domain like.
DR InterPro; IPR000047; HTH lambrpressor.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00031; HTHREPRESSR.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 388 AA; 42320 MW; 1627705405566A6C CRC64;

Query Match 15.4%; Score 214.5; DB 2; Length 388;
Best Local Similarity 27.7%; Pred. No. 2.3e-09;
Matches 84; Conservative 33; Mismatches 91; Indels 95; Gaps 14;

QY 3 SNLQGTFLNNTQLAQFSEMKAPMCQYQVNSFYKLSPPGLGQPLAAGTP----- 52
DB 69 SHLOASSPVRTPDVPSPLSLSAHQH-----PRLGCSGPGRYTMGCSA 116

QY 53 -----HGITDILSR---PVATPN-----SSLLSG-----YPHVAGGLSSQ 86
DB 117 QPAPASKDLKFGIDRILSABFDPKVRGNTLRDLTSLLSAGRTGAHAPHLO-----PSS 171

QY 87 GVYV-----GQVGSFSGKAGNEYPTRT-RNCWADTGDWGRSARPCGNTDPPLS-----DT 136
DB 172 GQYFTALEPGEASSVLGALHTHPRSTAQQQFQDT-----FFGYPVAVLTQDT 218

QY 137 I-----HKKHTRPTTGHQIFALEKTEQTKYLAGEPARLAYSLGMTESQVQVWFQNR 191
DB 219 MPQTYKRSRSRAVFSNLQKLEKFEIQKYVTKPDKQLAAMLGLTDAQVQVWFQNR 278

QY 192 RTKWRKKSALFSSSTP-----RAPGGASGDRAASEN-----EDDEYNKPLD-PDSD 237
DB 279 RMRWRHSKEAQAKEKEKEEVRSPGAQGRVDPNSLSRSEAESDSDSLDWA PSD 338

QY 238 DEK 240
DB 339 SER 341

RESULT 39
PAX6 DROME
ID PAX6 DROME STANDARD; PRT; 857 AA.
AC O18381; Q8ST91; Q8V4P5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Paired box protein Pax-6 (Eyeless protein).
GN Name=ey; Synonyms=pax6; ORENames=CG1464;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM EMBRYONIC), SEQUENCE OF 1-56 FROM N.A.
RP (ISOFORM LARVAL), FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
RP SPECIFICITY.
RC STRAIN=Oregon-R; TISSUE=Embryo, and Imaginal disks;
RX MEDLINE=94323757; PubMed=7914031;
RA Quiring R., Walldorf U., Kloter U., Gehring W.J.;
RT "Homology of the eyeless gene of Drosophila to the Small eye gene in
RT mice and Aniridia in humans.";

Query Match 15.4%; Score 214.5; DB 1; Length 857;
Best Local Similarity 27.0%; Pred. No. 6.2e-09;

Matches 64; Conservative 32; Mismatches 82; Indels 59; Gaps 6;

QY 12 NNTQLAQRSEKAPKQCVQSVQNSFKLS-----PPGLGQOLAAGTPHGTIDILSRVATPN 67

Db 307 NHQALQQHQOQSQWPPRHYS--GSWYPTSLSEIPISSAPNIASVTAYASGPSLAHSLSPN 364

QY 68 SLLSGYPHVAGFGGLSSGVVYQVQGSFSAKAGNEYPTTRNCWADT----- 115

Db 365 -----DIELSAIGHQ-----RNCPVATEDIHLKKELOGH 394

QY 116 -----GODWRGSRPCGNTDPLSDTIHKK--HTRPTFTGHQIPALEKTEPQTKY 164

Db 395 QSDTSGEGENSGASNGNIGTDDQARLILKRLQRNRTSFTNDQIDSLSEKFERHY 454

QY 165 LAGPERARLAYSIGMTESQVQVQFQNRRTYKWKKSALPSSSTPRAPGASGDRAAS 221

Db 455 PDVFARERLAGLIGLPEARIQVWFNRRAKWRREKLRNQRTPNSTGASATSSSTS 511

RESULT 40

DLX2 HUMAN

ID DLX2 HUMAN STANDARD; PRT; 328 AA.

AC Q07687;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Homeobox protein DLX-2.

GN Name=DLX2;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxId=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97001148; PubMed=8812481; DOI=10.1006/geno.1996.0387;

RA McGinniss T., Porteus M.H., Sniga S., Bulfone A., Kingsley C.,

RA Qiu M., Liu J.K., Long J.E., Xu D., Rubenstein J.L.R.,

RT "Sequence, organization, and transcription of the Dlx-1 and Dlx-2

RT locus.";

RL Genomics 35:473-485(1996).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.M.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywicki M.I., Skaleka U., Smallos D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RN [3]

RP SEQUENCE OF 85-328 FROM N.A.

RX MEDLINE=94040778; PubMed=7901126; DOI=10.1016/0378-1119(93)90212-L;

RA Seleki D.J., Thomas N.E., Coleman P.D., Rogers K.E.;

RT "The human brain homeogene, DLX-2: cDNA sequence and alignment with

RT the murine homologue.";

RL Gene 132:301-303(1993).

RN [4]

RP SEQUENCE OF 152-217 FROM N.A.

RC TISSUE=Embryo;

RX MEDLINE=94181570; PubMed=7907794;

RA Simeone A., Acampora D., Pannese M., D'Esposito M., Stornaiuolo A.,

RA Gullisano M., Mallamaci A., Kastury K., Druck T., Huebner K.,

RA Boncinelli E.;

RT "Cloning and characterization of two members of the vertebrate Dlx

RT gene family.";

RL Proc. Natl. Acad. Sci. U.S.A. 91:2250-2254(1994).

CC !- FUNCTION: Likely to play a regulatory role in the development of

CC the ventral forebrain. May play a role in craniofacial patterning

CC and morphogenesis.

CC !- SUBCELLULAR LOCATION: Nuclear (Potential).

CC !- SIMILARITY: Belongs to the distal-less homeobox family.

CC !- SIMILARITY: Contains 1 homeobox domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U51003; AAB40302.1; -

DR EMBL; BC032558; AAB32558.1; -

DR EMBL; L07919; AAL19663.1; -

DR PIR; B53495; B53495.

DR PIR; G02469; G02469.

DR HSSP; P13297; 1IG7.

DR TRANSFAC; T02407; -

DR Genew; HGNC:2915; DLX2.

DR H-InvDB; HIX0002596; -

DR MIM; 126255; -

DR GO; GO:0003700; F:transcription factor activity; TAS.

DR GO; GO:0007420; P:brain development; TAS.

DR InterPro; IPR001356; Homeobox.

DR InterPro; IPR009057; Homeoboxdomain_like.

DR InterPro; IPR000047; HTH_lambdarepressor.

DR Pfam; PF00046; Homeobox; 1.

DR PRINTS; PR00024; HOMEBOX.

DR PRINTS; PR00031; HTHREPRESSR.

DR ProDom; PR000010; Homeobox; 1.

DR SMART; SMO0389; HOX; 1.

DR PROSITE; PS00027; HOMEBOX 1; 1.

DR PROSITE; PS00071; HOMEBOX 2; 1.

KW Developmental protein; DNA-binding; Homeobox; Nuclear protein.

FT DOMAIN 41 46 Poly-Ser.

FT DOMAIN 75 81 Poly-Gly.

FT DNA BIND 152 211 Homeobox.

FT DOMAIN 251 259 Poly-Gly.

FT DOMAIN 309 315 Poly-His.

FT DOMAIN 316 319 Poly-Gly.

SQ SEQUENCE 328 AA; 34242 MW; BB6A077256F58022 CRC64;

Query Match 15.2%; Score 212.5; DB 1; Length 328;

Best Local Similarity 28.9%; Pred. No. 2.7e-09;

Matches 68; Conservative 42; Mismatches 86; Indels 39; Gaps 10;

QY 19 FSEKAPM--COYSVQNSFYKLSPPGLQPQLAAGTPHGITDILSRPVATP-----NS 68

Db 5 FDSLVDAMHSTQIAASSTYHQHQPPSGGAGPGNSSSSSLHHPQSPPTLPVSTATDS 64

QY 69 SILSGYPHVAGFGGLSSGVVYQVQGSF---SKAGNEYPTTRNCWADTG-----Q 117

Db 65 SYTYNQHPAGGG--GGSPVA-HMGSVQVQASGLNNVPYSKSSY-DLGYTAATSYA 120

QY 118 DWRGARPCGNTDP-----LSDTIHKKHTRPTFTGHQIPALEKTPQTKYLAGP 168

Db 121 PYGTSSSPANPEKEDELEPIRIVNGPKVKPRTIYSSPQLAALQRRFQKQYLAIP 180

QY 169 ERRLAYSIGMTESQVQVQFQNRRTWK--KSALEPSSSTPRAPGASGDRAAS 221

Db 181 ERAELASLGLTQTOVKIWFQNRRSKPKKMWKSGEIPSE---QHPGASASPPCAS 232

Search completed: April 13, 2005, 17:16:19
Job time : 255.264 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2005, 17:06:25 ; Search time 17.4054 Seconds
(without alignments)
439.204 Million cell updates/sec

Title: US-09-998-861-7
Perfect score: 102
Sequence: 1 XTKTPSVKDIILSLPEQXRAXGA 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1421835 seqs, 332370683 residues

Total number of hits satisfying chosen parameters: 1421835

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUBCOMB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	66.7	324	9	US-09-749-728B-9
2	68	66.7	324	14	US-10-177-390-26
3	60	58.8	273	16	US-10-654-102-87
4	54	52.9	281	16	US-10-257-765-4
5	54	52.9	301	16	US-10-257-765-2
6	47	45.1	786	15	US-10-369-493-9609
7	46.5	45.6	371	9	US-09-320-337-76
8	46.5	45.6	401	15	US-10-236-031B-26
9	46	45.1	41	16	US-10-437-963-108748
10	46	45.1	365	15	US-10-362-437-8
11	46	45.1	365	16	US-10-654-102-90
12	45	44.1	579	15	US-10-425-114-55742
13	44	43.1	162	15	US-10-424-599-206705

14	44	43.1	310	15	US-10-424-599-280169	Sequence 280169,
15	44	43.1	315	15	US-10-424-599-264401	Sequence 264401,
16	44	43.1	388	16	US-10-437-963-117151	Sequence 117151,
17	44	43.1	572	15	US-10-425-114-36880	Sequence 36880, A
18	44	43.1	635	16	US-10-437-963-192708	Sequence 192708,
19	44	43.1	732	15	US-10-369-493-3983	Sequence 3983, Ap
20	43	42.2	87	15	US-10-654-102-88	Sequence 88, Appl
21	43	42.2	348	15	US-10-424-599-269192	Sequence 269192,
22	43	42.2	434	15	US-10-424-599-263974	Sequence 263974,
23	43	42.2	1819	14	US-10-161-051-97	Sequence 97, Appl
24	42	41.2	56	15	US-10-424-599-214472	Sequence 214472,
25	42	41.2	117	16	US-10-767-701-43965	Sequence 43965, A
26	42	41.2	211	15	US-10-425-114-61380	Sequence 61380, A
27	42	41.2	236	16	US-10-716-293-99	Sequence 99, Appl
28	42	41.2	240	10	US-09-805-354-13	Sequence 269192,
29	42	41.2	240	11	US-09-758-493-13	Sequence 97, Appl
30	42	41.2	240	14	US-10-144-259-13	Sequence 214472,
31	42	41.2	313	16	US-10-437-963-176289	Sequence 43965, A
32	42	41.2	481	15	US-10-335-977-8870	Sequence 61380, A
33	42	41.2	486	15	US-10-335-977-8871	Sequence 99, Appl
34	42	41.2	762	16	US-10-437-963-149460	Sequence 13, Appl
35	42	41.2	784	15	US-10-369-493-19364	Sequence 13, Appl
36	42	41.2	798	13	US-10-072-841-30	Sequence 176289,
37	42	41.2	798	14	US-10-219-631-30	Sequence 8870, Ap
38	41	40.2	105	15	US-10-282-122A-65700	Sequence 19364, A
39	41	40.2	129	16	US-10-767-701-62341	Sequence 30, Appl
40	41	40.2	192	10	US-09-764-891-4432	Sequence 65700, A
41	41	40.2	367	15	US-10-362-437-1	Sequence 62341, A
42	41	40.2	367	16	US-10-654-102-89	Sequence 4432, Ap
43	41	40.2	423	16	US-10-437-963-128665	Sequence 1, Appl
44	41	40.2	529	15	US-10-437-963-114061	Sequence 89, Appl
45	41	40.2	648	15	US-10-425-114-58403	Sequence 128665,
						Sequence 114061,
						Sequence 58403, A

ALIGNMENTS

RESULT 1
US-09-749-728B-9
; Sequence 9, Application US/09749728B
; Patent No. US20020142457A1
; GENERAL INFORMATION:
; APPLICANT: Umezawa, Akihiro
; APPLICANT: Hata, Jun-ichi
; APPLICANT: Fukuda, Keiichi
; APPLICANT: Ogawa, Satoshi
; APPLICANT: Sakurada, Kazuhiro
; APPLICANT: Gojo, Satoshi
; APPLICANT: Yamada, Yoji
; TITLE OF INVENTION: THE CELL HAVING THE POTENTIALITY OF DIFFERENTIATION INTO CARDIOMY
; FILE REFERENCE: 00766.000043
; CURRENT APPLICATION NUMBER: US/09/749,728B
; CURRENT FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: H11-372826
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: PCT-JP00-01148
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: PCT-JP00-07741
; PRIOR FILING DATE: 2000-11-02
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver.2.0
; SEQ ID NO 9
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-749-728B-9

Query Match 66.7%; Score 68; DB 9; Length 324;
Best Local Similarity 63.6%; Pred. No. 0.0036;
Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 2 TKTTPSVKDIILSLPEQXRAXGA 23
| |||||:|:|:|

RESULT 7
US-09-320-337-76
; Sequence 76, Application US/09320337
; Patent No. US20010016352A1
; GENERAL INFORMATION:
; APPLICANT: Bohineki, Robert J.,
; APPLICANT: Whitsett, Jeffrey A.,
; TITLE OF INVENTION: Nucleic Acid Sequences Controlling
; TITLE OF INVENTION: Lung Cell - Specific Gene Expression
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM P160
; OPERATING SYSTEM: WINDOWS 95
; SOFTWARE: MS WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/320,337
; FILING DATE: 26-MAY-1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/442,809
; FILING DATE: 17-MAY-1995
; APPLICATION NUMBER: 08/245,356
; FILING DATE: 18-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliot M.
; REGISTRATION NUMBER: 24,025
; REFERENCE/DOCKET NUMBER: 271010-447
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 973-994-1700
; TELEFAX: 973-994-1744
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 371 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: human thyroid transcription factor-1
US-09-320-337-76
Query Match 45.6%; Score 46.5; DB 9; Length 371;
Best Local Similarity 60.0%; Pred. No. 22;
Matches 12; Conservative 1; Mismatches 6; Indels 1; Gaps 1;
QY 4 TPFSVKDILS-LPEQXRAG 22
Db 9 TPFSVSDILSPLSESYKKVG 28
RESULT 8
US-10-236-031B-26
; Sequence 26, Application US/10236031B
; Publication No. US20030219760A1
; GENERAL INFORMATION:
; APPLICANT: Gordon, Gavin J.
; APPLICANT: Jensen, Roderick V.
; APPLICANT: Gullans, Steven R.
; APPLICANT: Bueno, Raphael
; TITLE OF INVENTION: Diagnostic and Prognostic Tests
; FILE REFERENCE: B00801/70265 (JRV/JAV)
; CURRENT APPLICATION NUMBER: US/10/236,031B
; CURRENT FILING DATE: 2002-09-05
; PRIOR APPLICATION NUMBER: US 60/317,389

; PRIOR FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/407,431
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-236-031B-26
Query Match 45.6%; Score 46.5; DB 15; Length 401;
Best Local Similarity 60.0%; Pred. No. 24;
Matches 12; Conservative 1; Mismatches 6; Indels 1; Gaps 1;
QY 4 TPFSVKDILS-LPEQXRAG 22
Db 39 TPFSVSDILSPLSESYKKVG 58
RESULT 9
US-10-437-963-108748
; Sequence 108748, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 108748
; LENGTH: 41
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_12973C.1.pep
US-10-437-963-108748
Query Match 45.1%; Score 46; DB 16; Length 41;
Best Local Similarity 64.7%; Pred. No. 2;
Matches 11; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 3 KTFPSVKDILS-LPEQXR 19
Db 2 KTFPSVKDILS-LPEQXR 18
RESULT 10
US-10-362-437-8
; Sequence 8, Application US/10362437
; Publication No. US20040053210A1
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M.
; APPLICANT: Briscoe, James
; APPLICANT: Ericson, Johan
; APPLICANT: Rubenstein, John L.
; APPLICANT: Sander, Maïke
; TITLE OF INVENTION: GENETIC DEMONSTRATION OF REQUIREMENT FOR NKX6.1, NKX2.2 AND NKX6
; FILE REFERENCE: 62166-A-PCT-US
; CURRENT APPLICATION NUMBER: US/10/362,437
; CURRENT FILING DATE: 2003-08-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 8
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Rat
US-10-362-437-8

Query Match 45.1%; Score 46; DB 15; Length 365;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 4 TPFSVKDILSLPEQXRAXGA 23
DB 94 TPFGINDILSRPMPVASGA 113

RESULT 11
US-10-654-102-90
; Sequence 90, Application US/10654102
; Publication No. US20040132679A1
; GENERAL INFORMATION:
; APPLICANT: CHAN, LAWRENCE
; APPLICANT: KOJIMA, HIDEKO
; TITLE OF INVENTION: INDUCTION OF PANCREATIC ISLET FORMATION
; FILE REFERENCE: P02409US1
; CURRENT APPLICATION NUMBER: US/10/654,102
; CURRENT FILING DATE: 2003-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 90
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-654-102-90

Query Match 45.1%; Score 46; DB 16; Length 365;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 4 TPFSVKDILSLPEQXRAXGA 23
DB 94 TPFGINDILSRPMPVASGA 113

RESULT 12
US-10-425-114-55742
; Sequence 55742, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 55742
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73009H05_FLI.pep
US-10-425-114-55742

Query Match 44.1%; Score 45; DB 15; Length 579;
Best Local Similarity 52.4%; Pred. No. 66;
Matches 11; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 3 KTFPSVKDILSLPEQXRAXGA 23

DB 191 KTFPSVHELLNLVAQIRATGA 211

RESULT 13

US-10-424-599-206705
; Sequence 206705, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 206705
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_28682C.1.pep
US-10-424-599-206705

Query Match 43.1%; Score 44; DB 15; Length 162;
Best Local Similarity 57.1%; Pred. No. 22;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 TPFSVKDILSLPEQ 17
DB 142 SPFAGKDFSVPRQ 155

RESULT 14

US-10-424-599-280169
; Sequence 280169, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 280169
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_95015C.1.pep
US-10-424-599-280169

Query Match 43.1%; Score 44; DB 15; Length 310;
Best Local Similarity 57.1%; Pred. No. 47;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 TPFSVKDILSLPEQ 17
DB 290 SPFEGKDLFSLPRQ 303

RESULT 15

US-10-424-599-264401
; Sequence 264401, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:

GENERAL INFORMATION:
 APPLICANT: Liu, Jingdong
 APPLICANT: Zhou, Yihua
 APPLICANT: Kovalic, Steve
 APPLICANT: Screen, Steven
 APPLICANT: Tabaska, Jackie

GENERAL INFORMATION:
 APPLICANT: Liu, Jingdong
 APPLICANT: Zhou, Yihua
 APPLICANT: Kovalic, Steve
 APPLICANT: Screen, Steven
 APPLICANT: Tabaska, Jackie

FILE REFERENCE: 38-10(53052) B

TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)R

; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 3983
 ; LENGTH: 732
 ; TYPE: PRT
 ; ORGANISM: Neurospora crassa
 US-10-369-493-3983

Query Match 43.1%; Score 44; DB 15; Length 732;
 Best Local Similarity 50.0%; Pred. No. 1.3e+02;
 Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 TKTPTSVKDIILSLPEQ 17
 DB 374 TTTPTSTEDVISLDDQ 389

RESULT 20

US-10-654-102-88
 ; Sequence 88, Application US/10654102
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CHAN, LAWRENCE
 ; APPLICANT: KOJIMA, HIDEYO
 ; TITLE OF INVENTION: INDUCTION OF PANCREATIC ISLET FORMATION
 ; FILE REFERENCE: P02409US1
 ; CURRENT APPLICATION NUMBER: US/10/654,102
 ; CURRENT FILING DATE: 2003-09-03
 ; NUMBER OF SEQ ID NOS: 194
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 88
 ; LENGTH: 87
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-654-102-88

Query Match 42.2%; Score 43; DB 16; Length 87;
 Best Local Similarity 50.0%; Pred. No. 16;
 Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 6 FSVKDIILSLPEQXRAXGA 23
 DB 11 FKVKDILDPTDNDSDGS 28

RESULT 21

US-10-424-599-269192
 ; Sequence 269192, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 269192
 ; LENGTH: 348
 ; TYPE: PRT
 ; ORGANISM: Glycine max

; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)..(348)
 ; OTHER INFORMATION: unsure at all Xaa locations
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_85100C.1.pep

US-10-424-599-269192

Query Match 42.2%; Score 43; DB 15; Length 348;
 Best Local Similarity 50.0%; Pred. No. 80;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 KTFPSVKDIILSLPE 16
 DB 135 KTFPEIRDIILPLD 148

RESULT 22

US-10-424-599-263974
 ; Sequence 263974, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 263974
 ; LENGTH: 434
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)..(434)
 ; OTHER INFORMATION: unsure at all Xaa locations
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_80388C.1.pep
 US-10-424-599-263974

Query Match 42.2%; Score 43; DB 15; Length 434;
 Best Local Similarity 36.8%; Pred. No. 1e+02;
 Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 5 PFSVKDIILSLPEQXRAXGA 23
 DB 229 PYGLKDIISVPKRYKTWGS 247

RESULT 23

US-10-161-051-97
 ; Sequence 97, Application US/10161051
 ; Publication No. US20030152945A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Peter Deak
 ; APPLICANT: David M Glover
 ; APPLICANT: Carol Midgley
 ; TITLE OF INVENTION: Cell cycle progression proteins
 ; FILE REFERENCE: CGI-021CP
 ; CURRENT APPLICATION NUMBER: US/10/161,051
 ; CURRENT FILING DATE: 2002-05-30
 ; PRIOR APPLICATION NUMBER: GB 0007268.6
 ; PRIOR FILING DATE: 2000-03-24
 ; NUMBER OF SEQ ID NOS: 194
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 97
 ; LENGTH: 1819
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 US-10-161-051-97

Query Match 42.2%; Score 43; DB 14; Length 1819;
 Best Local Similarity 53.8%; Pred. No. 5.5e+02;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;


```
QY 4 TPFSVKDILSLPE 16
Db 1417 SPLSVQDLIQTFE 1429

RESULT 24
US-10-424-599-214472
; Sequence 214472, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 214472
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_35696C.1.pap
US-10-424-599-214472

Query Match 41.2%; Score 42; DB 15; Length 56;
Best Local Similarity 57.1%; Pred. No. 14;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 TPFSVKDILSLPEQ 17
Db 36 SPFLGKDLFSLPRQ 49

RESULT 25
US-10-767-701-43965
; Sequence 43965, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 43965
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C893_1.pap
US-10-767-701-43965

Query Match 41.2%; Score 42; DB 16; Length 117;
Best Local Similarity 42.9%; Pred. No. 34;
Matches 9; Conservative 5; Mismatches 3; Indels 4; Gaps 1;

QY 3 KTFPSVKDIL-----SLPEQXR 19
Db 15 ETPFALRGVLGSSSSYPEQAR 35

RESULT 26
US-10-425-114-61380
; Sequence 61380, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 61380
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3061-001-A12_FLI.pap
US-10-425-114-61380

Query Match 41.2%; Score 42; DB 15; Length 211;
Best Local Similarity 57.1%; Pred. No. 67;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 KTFPSVKDILSLPE 16
Db 123 KTFDEVDRDIFHLPD 136

RESULT 27
US-10-716-293-99
; Sequence 99, Application US/10716293
; Publication No. US20040127416A1
; GENERAL INFORMATION:
; APPLICANT: Arizona Board of Regents, acting for and on behalf of,
; APPLICANT: Arizona State University (ASU)
; APPLICANT: Massia, Stephen P.
; APPLICANT: Ehteshami, Gholam R.
; TITLE OF INVENTION: Bioselective bioconjugates for
; TITLE OF INVENTION: anti-inflammatory/immunosuppressant therapies
; FILE REFERENCE: 130588.00025
; CURRENT APPLICATION NUMBER: US/10/716,293
; CURRENT FILING DATE: 2003-11-17
; PRIOR APPLICATION NUMBER: 10/295,734
; PRIOR FILING DATE: 2002-11-15
; NUMBER OF SEQ ID NOS: 219
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 99
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Integrin
US-10-716-293-99

Query Match 41.2%; Score 42; DB 16; Length 236;
Best Local Similarity 72.7%; Pred. No. 76;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 TPFSVKDILSL 14
Db 95 TPFSYKNVLSL 105

RESULT 28
US-09-805-354-13
; Sequence 13, Application US/09805354
; Publication No. US20030078375A1
; GENERAL INFORMATION:
; APPLICANT: Arnaout, M. Amin
; APPLICANT: Li, Rui
; APPLICANT: Xiong, Jian-Ping
; TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF
```


; FILE REFERENCE: 00786-536001
; CURRENT APPLICATION NUMBER: US/09/805.354
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: US 09/758,493
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-805-354-13

Query Match 41.2%; Score 42; DB 10; Length 240;
Best Local Similarity 72.7%; Pred. No. 77;
Matches 8; Conservative 2; Mismatches 0; Gaps 0;

Qy 4 TPFSVKDILSL 14
Db 75 TPFSYKNVLSL 85

RESULT 29
US-09-758-493-13
; Sequence 13, Application US/09758493
; Publication No. US20040086935A1
; GENERAL INFORMATION:
; APPLICANT: Arnaout, M. Amin
; APPLICANT: Li, Rui
; APPLICANT: Xiong, Jian-Ping
; TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 00786-804001
; CURRENT APPLICATION NUMBER: US/09/758,493
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-758-493-13

Query Match 41.2%; Score 42; DB 11; Length 240;
Best Local Similarity 72.7%; Pred. No. 77;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 4 TPFSVKDILSL 14
Db 75 TPFSYKNVLSL 85

RESULT 30
US-10-144-259-13
; Sequence 13, Application US/10144259
; Publication No. US20030109691A1
; GENERAL INFORMATION:
; APPLICANT: Arnaout, M. Amin
; APPLICANT: Li, Rui
; APPLICANT: Xiong, Jian-Ping
; TITLE OF INVENTION: VARIANT INTEGRIN POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 00786-548001
; CURRENT APPLICATION NUMBER: US/10/144,259
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/758,493
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-259-13

Query Match 41.2%; Score 42; DB 14; Length 240;
Best Local Similarity 72.7%; Pred. No. 77;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 4 TPFSVKDILSL 14
Db 75 TPFSYKNVLSL 85

RESULT 31
US-10-437-963-176289
; Sequence 176289, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 176289
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_74051C.1.pep
US-10-437-963-176289

Query Match 41.2%; Score 42; DB 16; Length 313;
Best Local Similarity 61.5%; Pred. No. 1.1e+02;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 KTFPSVKDILSLP 15
Db 231 RTPFDSKDPLTLP 243

RESULT 32
US-10-335-977-8870
; Sequence 8870, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0


```

; INFORMATION FOR SEQ ID NO: 8871:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 486 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
;   ORGANISM: Helicobacter pylori
; FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (B) LOCATION 1...486
; SEQUENCE DESCRIPTION: SEQ ID NO: 8871:
US-10-335-977-8871

Query Match          41.2%; Score 42; DB 15; Length 486;
Best Local Similarity 38.9%; Pred. No. 1.8e+02;
Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 3 KTFPSVKDILSLPEQXRA 20
    ||| :||: ||| :||:
DB 406 KTFPINLEDLRSLEBEIKS 423

RESULT 34
US-10-437-963-149460
; Sequence 149460, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 149460
; LENGTH: 762
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_49791C.1.pap
US-10-437-963-149460

Query Match          41.2%; Score 42; DB 16; Length 762;
Best Local Similarity 64.3%; Pred. No. 3e+02;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 KTFPSVKDILSLPE 16
    ||| :||: ||| :||:
DB 435 KPPLSMKRIASLPE 448

RESULT 35
US-10-369-493-19364
; Sequence 19364, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B

```


Thu Apr 14 08:26:23 2005

us-09-998-861-7.rapb

Query Match
Best Local Similarity 41.2%; Score 42; DB 15; Length 784;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy

Db

7 SVKDILSLPEQXRA 20
||:|||||

660 SVQDITELPENLRA 673

RESULT 36

US-10-072-841-30
; Sequence 30, Application US/10072841
; Publication No. US20020164708A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Dean
; Pytela, Robert
; Quaranta, Vito
; TITLE OF INVENTION: A Novel Integrin Beta Subunit and Uses
; Thereof
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States of America
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/072,841
; FILING DATE: 06-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,215
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: P31 8717
; REFERENCE/DOCKET NUMBER: P31 8717
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-8949
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-10-072-841-30

Query Match
Best Local Similarity 41.2%; Score 42; DB 13; Length 798;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy

Db

4 TPFSVKDILSL 14
||:|||||

215 TPFSYKNVLSL 225

RESULT 37

US-10-219-631-30
; Sequence 30, Application US/10219631
; Publication No. US20030064471A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Dean
; Pytela, Robert
; Quaranta, Vito
; TITLE OF INVENTION: A Novel Integrin Beta Subunit and Uses
; Thereof
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States of America
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/219,631
; FILING DATE: 14-AUG-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,215
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: P31 8715
; REFERENCE/DOCKET NUMBER: P31 8717
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-10-219-631-30

Query Match
Best Local Similarity 41.2%; Score 42; DB 14; Length 798;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy

Db

4 TPFSVKDILSL 14
||:|||||

215 TPFSYKNVLSL 225

RESULT 38

US-10-282-122A-65700
; Sequence 65700, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.

Query Match
Best Local Similarity 72.7%; Pred. No. 3.1e+02;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy

Db

4 TPFSVKDILSL 14
||:|||||

215 TPFSYKNVLSL 225

APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-03-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 65700
LENGTH: 105
TYPE: PRT
ORGANISM: Neisseria meningitidis
US-10-282-122A-65700

Query Match 40.2%; Score 41; DB 15; Length 105;
Best Local Similarity 66.7%; Pred. No. 44;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 FSVKDIILSLPEQXRA 20
DB 61 FQDKDGLLLPEQIRA 75

RESULT 39
US-10-767-701-62341
Sequence 62341, Application US/10767701
Publication No. US20040172684A1
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53535)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 62341
LENGTH: 129
TYPE: PRT
ORGANISM: Sorghum bicolor
FEATURE:
OTHER INFORMATION: Clone ID: 18059734.pep
US-10-767-701-62341

Query Match 40.2%; Score 41; DB 16; Length 129;
Best Local Similarity 50.0%; Pred. No. 56;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 KTFPSVKDILSLPE 16
DB 46 KTFEIRDIFHLPD 59

RESULT 40

US-09-764-891-4432
Sequence 4432, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4432
LENGTH: 192
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (158)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (180)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (183)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (184)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (192)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-4432

Query Match 40.2%; Score 41; DB 10; Length 192;
Best Local Similarity 50.0%; Pred. No. 89;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 TKTPFSVKDILSLPEQ 17
DB 105 TATPAPIRQLLSRPER 120

Search completed: April 13, 2005, 17:22:38
Job time : 18.4054 secs

THIS PAGE BLANK (USPIO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2005, 17:03:40 ; Search time 68.1554 Seconds
(without alignments)
286.963 Million cell updates/sec

Title: US-09-998-861-13
Perfect score: 1396
Sequence: 1 MESNLQGTFLNNTQLAQS.....LLLRKRRAPFVSLGHSV 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pap:*
6: /cgn2_6/ptodata/1/iaa/backfiles.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	557	39.9	367	4	US-09-009-816-2
2	230	16.5	57	4	US-09-009-816-6
3	218	15.6	333	2	US-08-712-948-1
4	217	15.5	284	3	US-08-589-028-6
5	217	15.5	284	3	US-08-784-582-6
6	217	15.5	284	3	US-08-785-271-6
7	212.5	15.2	328	4	US-09-900-527-2
8	212.5	15.2	330	2	US-08-712-948-2
9	211.5	15.2	330	4	US-09-949-016-6621
10	211.5	15.2	334	4	US-09-949-016-7988
11	211	15.1	284	2	US-08-320-148B-2
12	211	15.1	284	3	US-09-031-898-2
13	208	14.9	300	3	US-09-162-524-1
14	206.5	14.8	217	4	US-09-949-016-6509
15	205.5	14.7	228	4	US-09-949-016-9719
16	202	14.5	86	4	US-09-009-816-8
17	200.5	14.4	349	3	US-09-162-524-3
18	198.5	14.2	264	4	US-09-949-016-6881
19	198.5	14.2	309	4	US-09-949-016-9244
20	195	14.0	283	1	US-08-583-672-2
21	195	14.0	283	2	US-08-202-044-2
22	195	14.0	283	3	US-08-751-344B-2
23	192	13.8	401	4	US-09-949-016-11737
24	191	13.7	289	4	US-09-976-594-945
25	191	13.7	289	4	US-09-949-016-10744
26	189	13.5	300	4	US-09-949-016-10794
27	188	13.5	214	4	US-09-949-016-9034

Query Match 39.9%; Score 557; DB 4; Length 367;
Best Local Similarity 41.7%; Pred. No. 2.3e-50;

ALIGNMENTS

RESULT 1

US-09-009-816-2
; Sequence 2, Application US/09009816
; Patent No. 6436667
; GENERAL INFORMATION:
; APPLICANT: German, Michael
; APPLICANT: Permutt, M. Alan
; APPLICANT: Inoue, Hiroshi
; TITLE OF INVENTION: Human Nkx-6.1 Polypeptide-Encoding
; TITLE OF INVENTION: Nucleotide Sequences
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,816
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Francis, Carol L
; REGISTRATION NUMBER: 36,513
; REFERENCE/DOCKET NUMBER: 9076/082CIP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3400
; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 367 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-09-009-816-2

28	188	13.5	214	4	US-09-949-016-9035	Sequence 9035, Ap
29	188	13.5	279	4	US-09-949-016-6403	Sequence 6403, Ap
30	188	13.5	311	4	US-09-949-016-10383	Sequence 10383, A
31	185.5	13.3	416	4	US-09-949-016-8977	Sequence 8977, Ap
32	185	13.3	314	4	US-09-949-016-6437	Sequence 6437, Ap
33	185	13.3	344	4	US-09-949-016-9889	Sequence 9889, Ap
34	184.5	13.2	301	3	US-09-095-117-6	Sequence 6, Appli
35	184.5	13.2	301	4	US-09-949-016-6178	Sequence 6178, Ap
36	184.5	13.2	304	3	US-09-095-117-8	Sequence 8, Appli
37	184.5	13.2	305	4	US-09-949-016-9718	Sequence 9718, Ap
38	181.5	13.0	224	4	US-09-949-016-11713	Sequence 11713, A
39	181.5	13.0	225	4	US-09-949-016-7007	Sequence 7007, Ap
40	180.5	12.9	287	3	US-09-031-962D-2	Sequence 2, Appli
41	180.5	12.9	287	4	US-09-949-016-9729	Sequence 9729, Ap
42	178.5	12.8	255	3	US-09-031-962D-4	Sequence 4, Appli
43	178	12.8	240	4	US-09-636-735A-2	Sequence 12, Appli
44	178	12.8	240	4	US-09-636-735A-12	Sequence 20, Appli
45	177.5	12.7	100	3	US-09-031-962D-20	

Matches 145; Conservative 27; Mismatches 74; Indels 102; Gaps 13;

Qy 1 MESNLOQTFLNNQTLA--QFSNMKAPM-----COYSVQNSFYKLSPP-- 41

Db 7 MEGTRQAFLLSSPPLAALHSMKAEKTPLYPAAYPPPLPAGPPSSSSSSSSSPPLGT 66

Qy 42 ----GLGP-----QLAAGTPHGHITDILSRP-----VATPNSS----- 69

Db 67 HNPGLKPPATGGLSSLGSPQQLSAATPHGINILSRPMPVASGAALPASGSGSSS 126

Qy 70 -----LLSGYHVHAGFCGLS-----SQGVYGPQ-- 93

Db 127 SSSSASASSAGAAAAAASAPAGLLAGLPR----FSSLSPPPPPGLYFSPSAAA 183

Qy 94 ---VGSFESKAGNEYTRNRCWADTGD--WRGSRPCGNTPDP--LSDTIHKKKTRP 145

Db 184 VAAVGRVPKPLAELPGRTPIFWPGVMSPPWRDLAC--TPHOGSILLDKDGRKTRP 241

Qy 146 TFTGHQIPALEKTEQTKYLAGPERARLAYSLGTMESQVWVFNRRTKWKKSALFPSS 205

Db 242 TFSGQIPALEKTEQTKYLAGPERARLAYSLGTMESQVWVFNRRTKWKKHAEMAT 301

Qy 206 STPRAPGAGCDRAASEN--BDEYNKPLDPDSDDEKIRLLLRHRAA 251

Db 302 AKKQDSETERLKGAASENEEDDYNKPLDPNSDDEKITQLLKKHKS 349

RESULT 2

US-09-009-816-6

; Sequence 6, Application US/09009816

; Patent No. 6436667

; GENERAL INFORMATION:

; APPLICANT: German, Michael

; APPLICANT: Permutt, M. Alan

; APPLICANT: Inoue, Hiroshi

; TITLE OF INVENTION: Human Nkx-6.1 Polypeptide-Encoding

; NUCLEOTIDE SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Bozicevic & Reed, LLP

; STREET: 285 Hamilton Ave, Suite 200

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94301

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/009,816

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Francis, Carol L

; REGISTRATION NUMBER: 36,513

; REFERENCE/DOCKET NUMBER: 9076/082CIP2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-327-3400

; TELEFAX: 650-327-3231

; TELEX:

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 57 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FRAGMENT TYPE: internal

US-09-009-816-6

Query Match 16.5%; Score 230; DB 4; Length 57;

Best Local Similarity 86.8%; Pred. No. 6e-17; Indels 0; Gaps 0;

Matches 46; Conservative 2; Mismatches 5;

Qy 133 LSDTIHKKKTRPTFTTGHQIPALEKTEQTKYLAGPERARLAYSLGTMESQVK 185

Db 5 LLDKGRKTRPTFTFSSQIFALEKTEQTKYLAGPERARLAYSLGTMESQVK 57

RESULT 3

US-08-712-948-1

; Sequence 1, Application US/08712948

; Patent No. 5850002

; GENERAL INFORMATION:

; APPLICANT: Korsemyer, Stanley J.

; TITLE OF INVENTION: HOX11 Gain and Loss of Function Murine

; NUCLEOTIDE SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Patrea L. Pabst

; STREET: 1100 Peachtree Street, Suite 2800

; CITY: Atlanta

; STATE: Georgia

; COUNTRY: USA

; ZIP: 30309-4530

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/712,948

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/231,728

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Pabst, Patrea L.

; REGISTRATION NUMBER: 31,284

; REFERENCE/DOCKET NUMBER: WU104

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (404)-815-6508

; TELEFAX: (404)-815-6555

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 333 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FRAGMENT TYPE: N-terminal

; ORIGINAL SOURCE:

; ORGANISM: Murine

US-08-712-948-1

Query Match 15.6%; Score 218; DB 2; Length 333;

Best Local Similarity 29.9%; Pred. No. 1.5e-14;

Matches 73; Conservative 29; Mismatches 78; Indels 64; Gaps 9;

Qy 36 YKLSPPGLGPQLAAGTPHGHITDILSRPVATNSLLSGYHVHAGFCG-----LSSQ 86

Db 74 YGAGGPG-GPGGPGAGGGGACSMGPLFSGYVNNMALAGPFGGCGGGGAGGAGALSA 132

Qy 87 GVV-----YGPQVGSFSK-----AGNEYPTTRNRCWADTGDW 119

Db 133 GVIRVPAHRPLAGAVAHPOPLATGLPTVPSPVAVGVNVLGLTFF-----WMSNRY 186

Qy 120 ---RGSARPCGNTPDPLSDTIHKKKTRPTFTTGHQIPALEKTEQTKYLAGPERARLAYS 176

Db 187 TKDRTHGHPYQNRTP-----KKKPRTSFTRLOICELEKHFHQKYLASAERAALAKA 240
 QY 177 LQMTSSQVQVWFQNRRTKWKKSALPSSSTPRAPGGASGDRAASENDEYVK-----230
 Db 241 LQMTDAQVQVWFQNRRTKWKKSALPSSSTPRAPGGASGDRAASENDEYVK-----230
 QY 231 PLDP 234
 Db 295 PADP 298

RESULT 4
 US-08-589-028-6
 ; Sequence 6, Application US/08589028
 ; Patent No. 6087129
 ; GENERAL INFORMATION:
 ; APPLICANT: Newgard, Christopher B.
 ; APPLICANT: Halban, Philippe A.
 ; APPLICANT: No. 6087129mington, Karl D.
 ; APPLICANT: Clark, Samuel A.
 ; APPLICANT: Thigpen, Anice E.
 ; APPLICANT: Quade, Christian
 ; APPLICANT: Kruse, Fred
 ; TITLE OF INVENTION: Recombinant Expression of Proteins From
 ; TITLE OF INVENTION: Secretary Cell Lines
 ; NUMBER OF SEQUENCES: 50
 ; CORRESPONDENCE ADDRESS:
 ; STREET: P. O. Box 4433
 ; CITY: Houston
 ; STATE: TX
 ; COUNTRY: USA
 ; ZIP: 77210-4433
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION NUMBER: US/08/589,028
 ; FILING DATE: Concurrently Herewith
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Highlander, Steven L.
 ; REGISTRATION NUMBER: 47,642
 ; REFERENCE/DOCKET NUMBER: UTSD:426\HYL
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (512) 418-3000
 ; TELEFAX: (512) 474-7577
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 284 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; US-08-589-028-6

Query Match 15.5%; Score 217; DB 3; Length 284;
 Best Local Similarity 30.3%; Pred. No. 1.6e-14;
 Matches 70; Conservative 26; Mismatches 95; Indels 40; Gaps 9;
 QY 25 PWCQSVQNSFYKSLPPGLGPQ-----LAAGTPHGITDILSRVPATPNSLLSGYPHV 77
 Db 33 PACLY-----MGRQPPPPPPQFTSSLSGSLGQSPDISPYEVPPLASDDPAGAHLLHHL 87
 QY 78 AGFGGLSSQGVYGVQVGSFSGAGNEYPTTR-----NCWADTGDWRGSRAPC 126
 Db 88 PAQLGLAHPPP--GPPNGTEPGGLEPNNRVQLPPPPMKSSTKAHAW--KGQ-WAGGA---139
 QY 127 GNTPPPLSDTIHKKHTRPTTGHQIFALEKTFEQTCKYLAGPERARLAYSGLGWTESQVKV 186
 Db 140 -YTAEP-----BENKRTRTAYTAQLLEKEKFLFNKYISRRPRRVELAVMLNLTERRHIKI 193

QY 187 WFORRTKWKKSALPSSSTPRAPGGA---SGDRAASENDEYVKPLDP 234
 Db 194 WFORNRMKWKEEDKKRSSGTSPGGGGEEPEQDCAVTSGRELLAVPLPP 244
 RESULT 5
 US-08-784-582-6
 ; Sequence 6, Application US/08784582
 ; Patent No. 6110707
 ; GENERAL INFORMATION:
 ; APPLICANT: Newgard, Christopher B.
 ; APPLICANT: Halban, Philippe A.
 ; APPLICANT: No. 6110707mington, Karl D.
 ; APPLICANT: Clark, Samuel A.
 ; APPLICANT: Thigpen, Anice E.
 ; APPLICANT: Quade, Christian
 ; APPLICANT: Kruse, Fred
 ; APPLICANT: McGarity, Dennis
 ; TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM
 ; TITLE OF INVENTION: SECRETORY CELL LINES
 ; NUMBER OF SEQUENCES: 79
 ; CORRESPONDENCE ADDRESS:
 ; STREET: P. O. Box 4433
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: USA
 ; ZIP: 77210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/784,582
 ; FILING DATE: Concurrently Herewith
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/028,427
 ; FILING DATE: 15-OCT-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/589,028
 ; FILING DATE: 19-JAN-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Highlander, Steven L.
 ; REGISTRATION NUMBER: 37,642
 ; REFERENCE/DOCKET NUMBER: UTSD:514
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 512/418-3000
 ; TELEFAX: 512/474-7577
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 284 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; US-08-784-582-6

Query Match 15.5%; Score 217; DB 3; Length 284;
 Best Local Similarity 30.3%; Pred. No. 1.6e-14;
 Matches 70; Conservative 26; Mismatches 95; Indels 40; Gaps 9;
 QY 25 PWCQSVQNSFYKSLPPGLGPQ-----LAAGTPHGITDILSRVPATPNSLLSGYPHV 77
 Db 33 PACLY-----MGRQPPPPPPQFTSSLSGSLGQSPDISPYEVPPLASDDPAGAHLLHHL 87
 QY 78 AGFGGLSSQGVYGVQVGSFSGAGNEYPTTR-----NCWADTGDWRGSRAPC 126
 Db 88 PAQLGLAHPPP--GPPNGTEPGGLEPNNRVQLPPPPMKSSTKAHAW--KGQ-WAGGA---139
 QY 127 GNTPPPLSDTIHKKHTRPTTGHQIFALEKTFEQTCKYLAGPERARLAYSGLGWTESQVKV 186

Db 140 -YTAEP-----EENKRTTAYTRAQLLEKEFLENKYISRRRVELAVMLNLTERHIKI 193
QY 187 WFNQRTKWRKKSALPSSSTPRAPGGA---SGDRAASENEDDEYNKPLDP 234
Db 194 WFNQRMWKKEEDKRSRSGTSPGGGGEPEQDCAVTSGEELLAVPPLPP 244

RESULT 6
US-08-785-271-6
; Sequence 6, Application US/08785271
; Patent No. 6194176
; GENERAL INFORMATION:
; APPLICANT: Newgard, Christopher B.
; APPLICANT: Halban, Philippe A.
; APPLICANT: No. 6194176minston, Karl D.
; APPLICANT: Clark, Samuel A.
; APPLICANT: Thigpen, Anice E.
; APPLICANT: Quade, Christian
; APPLICANT: Kruse, Fred
; TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM
; TITLE OF INVENTION: SECRETORY CELL LINES
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/785,271
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/589,028
; FILING DATE: 19-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UTSD:513
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 284 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-785-271-6

Query Match 15.5%; Score 217; DB 3; Length 284;
Best Local Similarity 30.3%; Pred. No. 1.6e-14;
Matches 70; Conservative 26; Mismatches 95; Indels 40; Gaps 9;

QY 25 PNCQYVNSQVYVGVQVSGFKAGNEYTRTR-----LAAGTPHGITDILSRPVATPNSLLSGYPHV 77
Db 33 PACLY-----MGRQPPPPPPQFTSSLSGLEQSPDIPVEVPLASDDPAGAHLLHL 87
QY 78 AGFGGLSQGVYVGVQVSGFKAGNEYTRTR-----NCWADTGQDWRGSRPC 126
Db 88 PAQGLAHPPP--GPPFNGTEPGGLEPNRVQLPFPNMKSTKAHAW--KGQ-WAGGA--- 139
QY 127 GNTDPLSDTIHKKHKTPTFTGHQIFALEKTEQTKYLAGPERARLAYSLGWTESQVKV 186
Db 140 -YTAEP-----EENKRTTAYTRAQLLEKEFLENKYISRRRVELAVMLNLTERHIKI 193
QY 187 WFNQRTKWRKKSALPSSSTPRAPGGA---SGDRAASENEDDEYNKPLDP 234

Db 194 WFNQRMWKKEEDKRSRSGTSPGGGGEPEQDCAVTSGEELLAVPPLPP 244

RESULT 7
US-09-900-527-2
; Sequence 2, Application US/09900527
; Patent No. 6602680
; GENERAL INFORMATION:
; APPLICANT: Rubenstein, John L.
; APPLICANT: Mione, Marina
; APPLICANT: Anderson, Stewart
; APPLICANT: Stuehmer, Thorsten
; APPLICANT: Yun, Kyuson
; TITLE OF INVENTION: Production of Gabaergic Cells
; FILE REFERENCE: UCSF184
; CURRENT APPLICATION NUMBER: US/09/900,527
; CURRENT FILING DATE: 2001-07-05
; PRIOR FILING DATE: 2001-07-14
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-900-527-2

Query Match 15.2%; Score 212.5; DB 4; Length 328;
Best Local Similarity 28.9%; Pred. No. 5.8e-14;
Matches 68; Conservative 42; Mismatches 86; Indels 39; Gaps 10;

QY 19 FSEMKAPM--COYSVQNSFYKLSPPGLPQLAAGTPHGITDILSRPVATP-----NS 68
Db 5 FDSLVADMHSQTIAASSTIHOHQPPGCGGAGPCGNSSSSSSLHKPQSPILPVSTATDS 64
QY 69 SLLSGYPHVAGFGLSQGVYVGVQVSGF-----SKAGNEYTRTRNCWADTG-----Q 117
Db 65 SYTINQHPAGGGG--GGGSPYA--HMGSYQYQASGLNNVPYSKSSY-DLGVTAAVTSVA 120
QY 118 DWGCSARPCGNTDPP-----LSDTIHKKHKTPTFTGHQIFALEKTEQTKYLAGP 168
Db 121 PYGTSSSPANNEPEKEDLEPIRVNGPKVKRPRTIYSSFQLAALQRRFQKTQYLLP 180
QY 169 ERARLAYSLGWTESQVKVFNQRTKWRK--KSALEPSSSTPRAPGGASGDRAAS 221
Db 181 ERAELASLGITQTVKIWFQNRKSKFKMKWSGEIPSE---QHPGASASPPCAS 232

RESULT 8
US-08-712-948-2
; Sequence 2, Application US/08712948
; Patent No. 5850002
; GENERAL INFORMATION:
; APPLICANT: Kormeyer, Stanley J.
; TITLE OF INVENTION: HOX11 Gain and Loss of Function Murine
; TITLE OF INVENTION: Models
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/712,948
; FILING DATE:

TITLE OF INVENTION: Nucleotide Sequences
 NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Bozicevic & Reed, LLP
 STREET: 285 Hamilton Ave, Suite 200
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94301
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/009,816
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Francis, Carol L
 REGISTRATION NUMBER: 36,513
 REFERENCE/DOCKET NUMBER: 9076/082CIP2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-327-3400
 TELEFAX: 650-327-3231
 TELEX:
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 86 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 US-09-009-816-8

Query Match 14.5%; Score 202; DB 4; Length 86;
 Best Local Similarity 57.4%; Pred. No. 1e-13;
 Matches 39; Conservative 11; Mismatches 16; Indels 2; Gaps 1;
 Qy 186 VWFQNRRTKWRKKALEPSSSTPRAPGASGDRAASEN--EDDEYNKPLDPSDDDEKIRL 243
 Db 1 VWFQNRRTKWRKKALEPSSSTPRAPGASGDRAASEN--EDDEYNKPLDPSDDDEKIRL 243
 Qy 244 LIRKRAA 251
 Db 61 LIRKRAA 251

RESULT 17
 US-09-162-524-3
 Sequence 3, Application US/09162524
 Patent No. 6387656
 GENERAL INFORMATION:
 APPLICANT: Jessell, Thomas M.
 APPLICANT: Tanabe, Yasuo
 APPLICANT: William, Christopher
 TITLE OF INVENTION: Gene Encoding MNR2 and Uses Thereof
 FILE REFERENCE: 57477/jpw/w1
 CURRENT APPLICATION NUMBER: US/09/162,524
 CURRENT FILING DATE: 1998-09-29
 NUMBER OF SEQ ID NOS: 4
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 3
 LENGTH: 349
 TYPE: PRT
 ORGANISM: chick embryo
 US-09-162-524-3

Query Match 14.4%; Score 200.5; DB 3; Length 349;

Best Local Similarity 25.4%; Pred. No. 1.2e-12;
 Matches 69; Conservative 29; Mismatches 73; Indels 101; Gaps 10;
 Qy 39 SPPGLGPO-----LAAGTGHGTTDILSRVATPNSLLSGYP-----HVAGFG 81
 Db 67 SPPRLPAHCALLPKAAFLGGGPG-----GHPQHIALGHPAGPG 108
 Qy 82 GLSQGVYVYGPQVGSFSGKAGNEYPTRTRNCWADTGDMRGSAAPCGNTDPL-----133
 Db 109 G---PGLYGHVYG-YPALGGHPALSYSYQVQ-----AHPAFSADPIKLSAGTFQ 158
 Qy 134 -----SDTIHKKHTRTFTTGHQIFALEKTFEOTKYLAP 168
 Db 159 LDOWLRASATAGMILPKMPDFGSAQSNLLGKCRPRPTAFTSQOLLEHOFKLNKYLSP 218
 Qy 169 ERAFLAYSLGWTSQVQVWFQNRRTKWR--KKSALPSSSTPRAPGASGD-----217
 Db 219 KRFEVATSLMLTETQVQVWFQNRRTKWR--KKSALPSSSTPRAPGASGD-----217
 Qy 218 -----RAASENEDDEYNKPLDPSDDDEK 240
 Db 279 LPPGKGGRRRLRLPDSE---PEDEEEEEE 307

RESULT 18

US-09-949-016-6881
 Sequence 6881, Application US/09949016
 Patent No. 6812339
 GENERAL INFORMATION:
 APPLICANT: VENTER, J. Craig et al.
 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 FILE REFERENCE: CL001307
 CURRENT APPLICATION NUMBER: US/09/949,016
 CURRENT FILING DATE: 2000-04-14
 PRIOR APPLICATION NUMBER: 60/241,755
 PRIOR FILING DATE: 2000-10-20
 PRIOR APPLICATION NUMBER: 60/237,768
 PRIOR FILING DATE: 2000-10-03
 PRIOR APPLICATION NUMBER: 60/231,498
 PRIOR FILING DATE: 2000-09-08
 NUMBER OF SEQ ID NOS: 207012
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 6881
 LENGTH: 264
 TYPE: PRT
 ORGANISM: Human
 US-09-949-016-6881

Query Match 14.2%; Score 198.5; DB 4; Length 264;
 Best Local Similarity 24.8%; Pred. No. 1.3e-12;
 Matches 67; Conservative 48; Mismatches 96; Indels 59; Gaps 13;
 Qy 5 LOGTFLLNNTOLAOFSEKAPMCOYSVQNSY-KLSPGLGQLAAGTGHGTTDILSRPV 63
 Db 2 IMSSYLM-----SNYIDPKFPCEYSQNSYIPEHSPEYYGRTRESGQHHELYPPPP 57
 Qy 64 ATPNSLLSGYPH-----VAGFGGLSSQGVYGPVGSFSGKAGNEYPTRTRNCWAD--114
 Db 58 PRPS-----YPERQYSCSTSLQPG--NSRG--HGP-----AQAGHHPEKSKICEPAP 102
 Qy 115 -TGQDWRGSAAP--CGN-TPD-----PLSDTIH-----KKHTRPT 146
 Db 103 LSGASASPSAPPACSQAPDHPSSAASKQPIVYPMKKIHVSTVNPYNGEPEKRSAA 162
 Qy 147 FTGHQIFALEKTFEOTKYLAPERARLAYSIGMTESQVQVWFQNRRTKWRKKSALPSSS 206
 Db 163 YTROQVLEKEBFHYNRYLTRRRIETIAHSLCLSERQIKIWFQNRRTKWRKKSALPSSS 222
 Qy 207 TPRAPGASGD--RAASENEDDEYNKPLDP 234
 Db 223 RSAPPAGAAPSTLSAATPGTSDHSQSATP 252

RESULT 19
US-09-949-016-9244
; Sequence 9244, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9244
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9244

Query Match 14.2%; Score 198.5; DB 4; Length 309;
Best Local Similarity 24.8%; Pred. No. 1.6e-12;
Matches 67; Conservative 48; Mismatches 96; Indels 59; Gaps 13;
QY 5 LOGTELLANTQAQSEMKAPMCQYVNSFY-KLSPGLGQPLAAGTGHGIDILSRPV 63
DB 47 IMSSYLM-----SNYIDPKFCEBYNSQSYIPEHSBYGRTRESGQHQQELYP 102
QY 64 ATPNSLLSGYPH-----VAGFGLSSQGVYGPQVGSFSAKAGNEYPTTRNCWAD-- 114
DB 103 PRPS-----YPERQYSCTSLOGPG--NSRG--HGP-----AQAGHHPEKSQSCEAP 147
QY 115 -TGQDWRGSRAP--CGN-TPD-----PLSDTIH-----KKHTRPT 146
DB 148 LSGASAPSPAPPACQAPDPHPSSAASKQPIVYPMWKKIHVSTVNPVNGEPEKRSRTA 207
QY 147 PTGHOIFALEKTFEOTKYLAGEPARLAYSLGTMESQVKKVWFQNRRTKWKKSALPSSS 206
DB 208 YTRQVLEKEFHFYHNYLRRRRIEIAHSLCLSERQIKWPNRRMKWKDHLPLNTKV 267
QY 207 TPAPGGASGD--RAASENDEYKNPLDP 234
DB 268 RSAPPAGAAPSTLSAATGCTSEDHSQSATP 297

RESULT 20
US-08-583-672-2
; Sequence 2, Application US/08583672
; Patent No. 5741673
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R.
; APPLICANT: Leonard, James N.
; TITLE OF INVENTION: A NOVEL HOMEOBOX FACTOR THAT STIMULATES
; INSULIN EXPRESSION IN PANCREATIC ISLET CELLS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/583,672
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/106,936
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9422
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 283 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-583-672-2

Query Match 14.0%; Score 195; DB 1; Length 283;
Best Local Similarity 29.9%; Pred. No. 3.3e-12;
Matches 66; Conservative 23; Mismatches 86; Indels 46; Gaps 11;
QY 25 PMCYVNSQVNSFYKLSPPGLGQPLAA-----GTPHGITDILSRVA-TPNSSLLSGYPH 76
DB 33 PACLY-----MGRQPPPPPPQFAGSLGTLEQSGPPDISPYEVPPLADDPAGAH--HHH 85
QY 77 VAGFGLSSQGVYGPQVGSFSAKAGNEYPTTR-----NCWADTGQDWRGSRAP 125
DB 86 LPAQGLAHAPP--GPPFNGTETGGLPEPSRVHLPPPMKSTKAHAWK---SQWAGGA-- 138
QY 126 CGNTPDPLSDTIHKCKTTRPTTGHQIFALEKTFEOTKYLAGEPARLAYSLGTMESQV 185
DB 139 --YAEPP-----ENKTRTRAYTRAQLELEKEFLFNKYISRRPRVELAVMLNTERIK 191
QY 186 VWFQNRRTKWKKSALPSSSTPRAPGASGDRAASENEDD 226
DB 192 IWFQNRRTKWKKSALPSSSTPRAPGASGDRAASENEDD 226

RESULT 21
US-08-202-044-2
; Sequence 2, Application US/08202044
; Patent No. 5858973
; GENERAL INFORMATION:
; APPLICANT: Habener M.D., Joel F.
; APPLICANT: Miller Ph.D., Christopher P.
; TITLE OF INVENTION: NOVEL TRANSCRIPTION FACTOR AND USES
; THEREFOR
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,044
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Ph.D., Kathleen A.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: MGH-124XX

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 542-2290
 TELEFAX: (617) 451-0313
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 283 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-202-044-2

Query Match 14.0%; Score 195; DB 2; Length 283;
 Best Local Similarity 29.9%; Pred. No. 3.3e-12;
 Matches 66; Conservative 23; Mismatches 86; Indels 46; Gaps 11;
 QY 25 PMCOYSVQNSFYKLSPPGLGPOLAA-----GTPHGITDILSRPVA--TPNSSLLSGYPH 76
 DB 33 PACLY-----MGRQPPPPPTQFAGSLGTLEQSGSPDISPYEVPLADDPAGAHLL--HHH 85
 QY 77 VAGFGGLSSQGVYVGPVGSFSGKAGNEYPTTR-----NCWADTGDQWRGSGARP 125
 DB 86 LPAQLGLAHP--GPPNGTETGLESPSRVHLPPFWMKSTKAHAWK---SQWAGGA-- 138
 QY 126 CNTDPDPLSDTHKKHTRPTFTGHQIFALEKTFEQTLYLAGPERARLAYSLGWTESQVK 185
 DB 139 --YAAEP-----EENKTRTAYTRAQLELEKEFLFNKYISRPVRVELAVMLNLTERRHIK 191
 QY 186 VWFONRTKWKRSKSALEPSSSTPRAPGASGDRAASENEDD 226
 DB 192 IWFQNRMKWKKEEDKRRSGT--TSGGGGGE-----EPEQD 226

RESULT 22
 US-08-751-344B-2
 ; Sequence 2, Application US/08751344B
 ; Patent No. 6210960
 ; GENERAL INFORMATION:
 ; APPLICANT: Habener M.D., Joel F.
 ; APPLICANT: Miller Ph.D., Christopher P.
 ; TITLE OF INVENTION: NOVEL TRANSCRIPTION FACTOR AND USES
 ; TITLE OF INVENTION: THEREFOR
 ; NUMBER OF SEQUENCES: 29
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Banner & Witcoff, Ltd.
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: US
 ; ZIP: 02111
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WordPerfect 6.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/751,344B
 ; FILING DATE: 19-NO. 6210960-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/202,044
 ; FILING DATE: 23-Feb-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Williams Ph.D., Kathleen M.
 ; REGISTRATION NUMBER: 34,380
 ; REFERENCE/DOCKET NUMBER: 96,137-A (11274/02148)
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 345-9100
 ; TELEFAX: (617) 345-9111
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 283 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein

US-08-751-344B-2

Query Match 14.0%; Score 195; DB 3; Length 283;
 Best Local Similarity 29.9%; Pred. No. 3.3e-12;
 Matches 66; Conservative 23; Mismatches 86; Indels 46; Gaps 11;
 QY 25 PMCOYSVQNSFYKLSPPGLGPOLAA-----GTPHGITDILSRPVA--TPNSSLLSGYPH 76
 DB 33 PACLY-----MGRQPPPPPTQFAGSLGTLEQSGSPDISPYEVPLADDPAGAHLL--HHH 85
 QY 77 VAGFGGLSSQGVYVGPVGSFSGKAGNEYPTTR-----NCWADTGDQWRGSGARP 125
 DB 86 LPAQLGLAHP--GPPNGTETGLESPSRVHLPPFWMKSTKAHAWK---SQWAGGA-- 138
 QY 126 CGNTPDPLSDTHKKHTRPTFTGHQIFALEKTFEQTLYLAGPERARLAYSLGWTESQVK 185
 DB 139 --YAAEP-----EENKTRTAYTRAQLELEKEFLFNKYISRPVRVELAVMLNLTERRHIK 191
 QY 186 VWFONRTKWKRSKSALEPSSSTPRAPGASGDRAASENEDD 226
 DB 192 IWFQNRMKWKKEEDKRRSGT--TSGGGGGE-----EPEQD 226

RESULT 23

US-09-949-016-11737
 ; Sequence 11737, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; PRIOR FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 11737
 ; LENGTH: 401
 ; TYPE: PRT
 ; ORGANISM: Human
 ; US-09-949-016-11737

Query Match 13.8%; Score 192; DB 4; Length 401;
 Best Local Similarity 32.0%; Pred. No. 1.2e-11;
 Matches 74; Conservative 20; Mismatches 91; Indels 46; Gaps 12;
 QY 41 PGLGPOLAAAGTPHGITDILSRPVATPNSLLSGYPHVAGFGGLSSQG---VYVGPVGSF 97
 DB 143 PG-GAQQGAGLP-AQAALYGHV-----YGSAAAAAALAGQHPALSYSYPVQGA 192
 QY 98 SKAGNEYPTTRNCWADTQ--DW-----RGSARPCGNTPD-----PLSDTHKKHTRPT 146
 DB 193 HPA---HPADPIKLAGTTFOLDQWLRASTAGMILP--KMPDENSQAQSNLLGKCRPRTA 247
 QY 147 FTGHQIFALEKTFEQTLYLAGPERARLAYSLGWTESQVKWFOFNRRTKW-RKKSALPES- 204
 DB 248 FTSQQLLEHQPFLNKYLSRPKEFVATSLMLTETQVKIWFQNRMKWKESKAKAEOAA 307
 QY 205 --SSTPRAPGASGDRAASENEDDDEYNKPLDP-----DSDEK 240
 DB 308 QEREKQGGGGGAGKGAEEFPCABELLGPAPGKSGRRRLDLRSDPDE 358

RESULT 24

US-09-976-594-945
 ; Sequence 945, Application US/09976594
 ; Patent No. 6673549


```

; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 945
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 1505790CD1
US-09-976-594-945

Query Match 13.7%; Score 191; DB 4; Length 289;
Best Local Similarity 28.3%; Pred. No. 9e-12;
Matches 64; Conservative 30; Mismatches 76; Indels 56; Gaps 9;

Qy 20 SEMKAPMCQYSVQNSFYKLSPPGLGQPLAAGTTPHGTIDILSRPVATPNSLLSGYPH--- 76
Db 31 SOESTLPSSATSDSY-YSPTG-----GAPHGYS-----PTSASYGKALNPYQYQH 78

Qy 77 -VAGFGLSSQGVY----YGPQVGSFSGKAGNEYPTTRTNCWADTGDQWRGSGARPCGNTPD 131
Db 79 GVNGSAGYPKAKAYADYSYASYHQYGAYNRVPSATN-----OPEKYTE 124

Qy 132 P-----LSDTIHKKHTRPTFTTGHQIFALEKTFEOTKYLAGPERARLAYSLGWTESQVKW 187
Db 125 PEVRMVGKPKKVRKPRTIYSSFQALQRRFQKTYLALPERAELASGLTQTQVKIW 184

Qy 188 FQNRRTKWRK-----KSALEPSSSTPRA-----PGGAS 215
Db 185 FONKSKIKTKMKGEMPPHSPSSDDPMACNSQSPAVWEPQGSS 230

; RESULT 26
US-09-949-016-10794
; Sequence 10794, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10794
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10794

Query Match 13.5%; Score 189; DB 4; Length 300;
Best Local Similarity 28.3%; Pred. No. 1.6e-11;
Matches 58; Conservative 28; Mismatches 69; Indels 50; Gaps 6;

Qy 78 AGFGLSSQGVYGPQVGSFSGKAGNEYPTTRTNCWADT-GDQWRGSGARPC-----G 127
Db 39 SCAGDLRADGGGHSPTVAGTSASS--PAGSREGADSDGQPGCEADHCRRLVRDAG 96

Qy 128 NTPD---PLSDTIHKKHTRPTFTTGHQIFALEKTFEOTKYLAGPERARLAYSLGWTESQV 184
Db 97 TIREIVLPKGLDLDPRKRTFTSFTAEQYLRLEMEFQRCQYVVGRTTELARQLNLSQTV 156

Qy 185 KWVFQNRRTKWRKSA--LEPSSS-----TPR 209
Db 157 KWVFQNRRTKQKQDQSRDLKRRASSASEAFATSNILRLSQGLLSVPRAPSLALTPS 216

Qy 210 APGGASGDRAASENEDDEYNKPLDP 234
Db 217 LPGLPASHRGTSGLGDPNRNSSLNLP 241

; RESULT 27
US-09-949-016-9034
; Sequence 9034, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20

```

```

; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 945
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 1505790CD1
US-09-976-594-945

Query Match 13.7%; Score 191; DB 4; Length 289;
Best Local Similarity 28.3%; Pred. No. 9e-12;
Matches 64; Conservative 30; Mismatches 76; Indels 56; Gaps 9;

Qy 20 SEMKAPMCQYSVQNSFYKLSPPGLGQPLAAGTTPHGTIDILSRPVATPNSLLSGYPH--- 76
Db 31 SOESTLPSSATSDSY-YSPTG-----GAPHGYS-----PTSASYGKALNPYQYQH 78

Qy 77 -VAGFGLSSQGVY----YGPQVGSFSGKAGNEYPTTRTNCWADTGDQWRGSGARPCGNTPD 131
Db 79 GVNGSAGYPKAKAYADYSYASYHQYGAYNRVPSATN-----OPEKYTE 124

Qy 132 P-----LSDTIHKKHTRPTFTTGHQIFALEKTFEOTKYLAGPERARLAYSLGWTESQVKW 187
Db 125 PEVRMVGKPKKVRKPRTIYSSFQALQRRFQKTYLALPERAELASGLTQTQVKIW 184

Qy 188 FQNRRTKWRK-----KSALEPSSSTPRA-----PGGAS 215
Db 185 FONKSKIKTKMKGEMPPHSPSSDDPMACNSQSPAVWEPQGSS 230

; RESULT 25
US-09-949-016-10744
; Sequence 10744, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10744
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10744

Query Match 13.7%; Score 191; DB 4; Length 289;
Best Local Similarity 28.3%; Pred. No. 9e-12;
Matches 64; Conservative 30; Mismatches 76; Indels 56; Gaps 9;

Qy 20 SEMKAPMCQYSVQNSFYKLSPPGLGQPLAAGTTPHGTIDILSRPVATPNSLLSGYPH--- 76
Db 31 SOESTLPSSATSDSY-YSPTG-----GAPHGYS-----PTSASYGKALNPYQYQH 78

```


; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9034
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9034

Query Match 13.5%; Score 188; DB 4; Length 214;
Best Local Similarity 38.0%; Pred. No. 1.2e-11;
Matches 57; Conservative 16; Mismatches 67; Indels 10; Gaps 6;

QY 90 YGQVGSFSGKAGNEYP--TTRNCWADTGDWRGSAAPCGNTDPDLSDTIHKKHTRPTFT 148
Db ||||| -GGF--GGF--GGPLFPFFRTVNDYTHALLRHDPLGKLLWSPF-LQRLHKKRGQVRF 89

QY 149 GHQIFALEKTFTQTKYLAGPERARLAYSLGMTESQVQVWFQNRRTKWRKKSALPSSSTP 208
Db ||||| -GGF--GGF--GGPLFPFFRTVNDYTHALLRHDPLGKLLWSPF-LQRLHKKRGQVRF 89

QY 209 RAPGASGDRAASENED--DEYNKPLDPDS 236
Db 148 KKEELESLSDDSCDQRLPSEQNKGLSDS 177

RESULT 28
US-09-949-016-9035
; Sequence 9035, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9035
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9035

Query Match 13.5%; Score 188; DB 4; Length 214;
Best Local Similarity 38.0%; Pred. No. 1.2e-11;
Matches 57; Conservative 16; Mismatches 67; Indels 10; Gaps 6;

QY 90 YGQVGSFSGKAGNEYP--TTRNCWADTGDWRGSAAPCGNTDPDLSDTIHKKHTRPTFT 148
Db ||||| -GGF--GGF--GGPLFPFFRTVNDYTHALLRHDPLGKLLWSPF-LQRLHKKRGQVRF 89

QY 149 GHQIFALEKTFTQTKYLAGPERARLAYSLGMTESQVQVWFQNRRTKWRKKSALPSSSTP 208
Db ||||| -GGF--GGF--GGPLFPFFRTVNDYTHALLRHDPLGKLLWSPF-LQRLHKKRGQVRF 89

QY 209 RAPGASGDRAASENED--DEYNKPLDPDS 236
Db 148 KKEELESLSDDSCDQRLPSEQNKGLSDS 177

RESULT 29
US-09-949-016-6403

; Sequence 6403, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6403
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6403

Query Match 13.5%; Score 188; DB 4; Length 279;
Best Local Similarity 26.1%; Pred. No. 1.8e-11;
Matches 69; Conservative 35; Mismatches 86; Indels 74; Gaps 11;

QY 8 TELLNNTQLAQFSEMKAPMCQYSVQNSFYKLSPPGLGPQL-----AAGTP----HG 54
Db ||||| -LKTCDY-----FEKLSYVCPSLVVRPKPLHSCCTGSPSLRAYP 71

QY 55 ITDILSRPVATPNSLLSGYPHVAGFGLSSQGVY-----GPOVGSFSGKAGNEYPTR 107
Db ||||| -LKTCDY-----FEKLSYVCPSLVVRPKPLHSCCTGSPSLRAYP 71

QY 72 LLSVITRQ-PTVISHLVPAATPGIA--QALSCHQVTEAVSAEAPGAEALASSETEQPTP 128
Db ||||| -LKTCDY-----FEKLSYVCPSLVVRPKPLHSCCTGSPSLRAYP 71

QY 108 TRNCWADTGDWRGSAAPCGNTDPDLSDTIHKKHTRPTFTGHIQFALEKTFTQTKYLAG 167
Db ||||| -LKTCDY-----FEKLSYVCPSLVVRPKPLHSCCTGSPSLRAYP 71

QY 129 RQK-----KPRRSRTIETELQMLGLEKKFKQKYLST 160
Db ||||| -LKTCDY-----FEKLSYVCPSLVVRPKPLHSCCTGSPSLRAYP 71

QY 168 PERARLAYSLGMTESQVQVWFQNRRTKWRKKSALPSSSTPRAPGG-----ASGDRA 219
Db ||||| -LKTCDY-----FEKLSYVCPSLVVRPKPLHSCCTGSPSLRAYP 71

QY 161 PDRDLAQSLGLTQVKTWYQNRMRKWK-KWVLKGGQEAFTKGRPKKNSIPTSEIE 219
Db ||||| -LKTCDY-----FEKLSYVCPSLVVRPKPLHSCCTGSPSLRAYP 71

QY 220 ASE--NEDEYNKPLDPDSDEKI 241
Db ||||| -LKTCDY-----FEKLSYVCPSLVVRPKPLHSCCTGSPSLRAYP 71

QY 220 AEEKWNSQAQGEQLEPSQGEEL 243
Db ||||| -LKTCDY-----FEKLSYVCPSLVVRPKPLHSCCTGSPSLRAYP 71

RESULT 30
US-09-949-016-10383
; Sequence 10383, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10383
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10383


```
Query Match      13.5%; Score 188; DB 4; Length 311;
Best Local Similarity 26.1%; Pred. No. 2.1e-11;
Matches 69; Conservative 35; Mismatches 86; Indels 74; Gaps 11;

Qy 8 TPLNNTQLAQFSEKAMPQCVSVQNSFYKLSPPGLGPGQL-----AACTP-----HG 54
Db 56 TFWIDEI-----LSKETCDY-----PEKLSYVCSLSLVVRPKPLHSCCTGSPSLRAYP 103
Qy 55 ITDILSRVATPNSLLSGYPHVAGFGLSGGVY-----GPQVGSFKAGNEYPT 107
Db 104 LLSVITRQ-PTVISHLVATPGIA--QALSCHQVTEAVSAEAPGGEALASSESETEQPTP 160
Qy 108 TRNCWADTQDWRGARSARCGNTPDPLSDTIHKKHTRPTFGHQIFALEKTFEQTKYLAG 167
Db 161 RQK-----KPRSRTIFTELQMLGLEKFKQKOKYLST 192
Qy 168 PERARLAYSGLMTEQVWFONRRTKWRKKSALPSSSTPRAPCG-----ASGDRA 219
Db 193 PDRDLAQSGLTQVKTWTQNRMMK-KMWLKGQGEAPTYPKGRPKQNSIPTSEIE 251
Qy 220 ASE--NEDDEYNKPLDPDSDEKI 241
Db 252 AEEKWNSQAQGEQLERPSQGEEL 275

RESULT 31
US-09-949-016-8977
; Sequence 8977, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8977
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8977

Query Match      13.3%; Score 185.5; DB 4; Length 416;
Best Local Similarity 30.5%; Pred. No. 5.9e-11;
Matches 61; Conservative 26; Mismatches 72; Indels 41; Gaps 10;

Qy 40 PPGGLPQLAAGTPHGIITILSRVATPNSLLSGYPHVAGFGLSGGVYGPQVGSFSK 99
Db 91 PPGLNSEQQPQPPE-----PPPTLPSPSPTNP-----GGVPAKPKGPNASSSSA 138
Qy 100 AGNE--YP-----TRNCWADTQDWRGARSARCGNTPDPLSDTIHKKHTRPTTG 149
Db 139 TISKQIFPMKESRQNSKQKNSCATAGESCEDK-----SPPGPAS-----KRVRTA 187
Qy 150 HQIFALEKTFEQTKYLAGPERARLAYSGLMTEQVWFONRRTKWRK-----KSAL-BPS 204
Db 188 AQLVELEKEFHFNRYLCRRVEMANLNLTERQIKIWFQNERMYKKQAKGILHSPA 247
Qy 205 SSTP-RAP--GAASGDRAAS 221
Db 248 SQSPERSPLGGAAGHAVYS 267

RESULT 32
US-09-949-016-6437
```

```
; Sequence 6437, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6437
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6437

Query Match      13.3%; Score 185; DB 4; Length 314;
Best Local Similarity 34.6%; Pred. No. 4.4e-11;
Matches 62; Conservative 27; Mismatches 72; Indels 18; Gaps 9;

Qy 68 SLLSGYPHVAGFGLS--SQGVYGPQVGSFSKAGNEYPTRT-RNCWADTQDWRGARSARP 125
Db 24 SSLASAY---ADFSSCSQASGFQYNPIRTTFG-ATSGCPSLTGSCSLGTLRDHSS--P 77
Qy 126 CGNTPDPLSDTIH-----KKKHTRPTFGHQIFALEKTFEQTKYLAGPERARLAYSIG 178
Db 78 YAAVPYKLF-TDHGGLNEKRRQIRITFTTSAQLKELERVFETHYPIYTRTELAKID 136
Qy 179 MTESQVWVFONRRTKWRKKSALPSSSTPRAPCGASGDRAASENEDDEYN-KPLDPPDS 236
Db 137 LTEARVQVWFQNRRAKFRKQER-AAAAAAAANKGSSGKSSDRDDESKAKSTDPDS 194

RESULT 33
US-09-949-016-9889
; Sequence 9889, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9889
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9889

Query Match      13.3%; Score 185; DB 4; Length 344;
Best Local Similarity 34.6%; Pred. No. 5.1e-11;
Matches 62; Conservative 27; Mismatches 72; Indels 18; Gaps 9;

Qy 68 SLLSGYPHVAGFGLS--SQGVYGPQVGSFSKAGNEYPTRT-RNCWADTQDWRGARSARP 125
Db 54 SSLASAY---ADFSSCSQASGFQYNPIRTTFG-ATSGCPSLTGSCSLGTLRDHSS--P 107
```


QY 126 CGNTDPLSDTH-----KKHTRPTFTGHQIFALEKTFBQTKYLAPERARLAYSIG 178
Db 108 YAAVPYKLF-TDGGLINEKRRKORITFTTSQKLKELERVAETHYDPDIYEEELAKID 166
QY 179 MTESQVWVFQNRRTKWRKSALEPSSSTPRAPGASGDRAASENEDDEYN-KPLDPOS 236
Db 167 LTEARVQVWFQNRRAFRKQER-AAAAAAAAGKSSGKSDSRDDESKEAKSTDPOS 224

RESULT 34
US-09-095-117-6
; Sequence 6, Application US/09095117
; Patent No. 6228582
; GENERAL INFORMATION:
; APPLICANT: Rodier, Patricia M.
; APPLICANT: Ingram, Jennifer L.
; APPLICANT: Figlewicz, Denise A.
; APPLICANT: Hyman, Susan L.
; APPLICANT: Stodgell, Christopher J.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS WHICH ARE
; TITLE OF INVENTION: ASSOCIATED WITH AUTISM SPECTRUM DISORDERS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603-1051
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/095,117
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/049,803
; FILING DATE: 17-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 176/60181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 301 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-095-117-6

Query Match 13.2%; Score 184.5; DB 3; Length 301;
Best Local Similarity 26.6%; Pred. No. 4.7e-11;
Matches 66; Conservative 25; Mismatches 112; Indels 45; Gaps 9;
QY 22 MKAPMCQVSNQSFY--KLSPPLGQPLAAGTPHGTIDILSRVATPN-----SSLSS 72
Db 52 LSSPAFQ---QNSGYPAQPPSTLGVPPFAPSAPYAPACSPSYGSPYPLQSGEGD 108
QY 73 GYPHVAGFG-----GLSSQGVYGVQVGSFSGKAGNEYTRTRNCWADTGQDWRGSAR 124
Db 109 GYFHPSSVGAQLGSLSDYGAGGAGPGYPQHPYQNEQTASAPAYADLLSE--DKET 166
QY 125 PCGNTPD-PLSDTH--KKHTRP-----TFTGHQIFALEKTFEOTK 163
Db 167 PCPSEPNTPTARTFDMVKRNPPKTKVSBPGLSPSGLTNTFTTQLTELEKEFHFNK 226

QY 164 YLAGPERARLAYSIGMTESQVWVFQNRRTKWRKSALEPSSSTPRAPGASGDRAASEN 223
Db 227 YLSRARRVEIAATLEINETQVKINFORNMKQKRE--REGGRVPPAPPGCPKEAAGDAS 284
QY 224 EDDEYNKP 231
Db 285 DQSTCTSP 292

RESULT 35
US-09-949-016-6178
; Sequence 6178, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6178
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-6178

Query Match 13.2%; Score 184.5; DB 4; Length 301;
Best Local Similarity 26.6%; Pred. No. 4.7e-11;
Matches 66; Conservative 25; Mismatches 112; Indels 45; Gaps 9;

QY 22 MKAPMCQVSNQSFY--KLSPPLGQPLAAGTPHGTIDILSRVATPN-----SSLSS 72
Db 52 LSSPAFQ---QNSGYPAQPPSTLGVPPFAPSAPYAPACSPSYGSPYPLQSGEGD 108
QY 73 GYPHVAGFG-----GLSSQGVYGVQVGSFSGKAGNEYTRTRNCWADTGQDWRGSAR 124
Db 109 GYFHPSSVGAQLGSLSDYGAGGAGPGYPQHPYQNEQTASAPAYADLLSE--DKET 166
QY 125 PCGNTPD-PLSDTH--KKHTRP-----TFTGHQIFALEKTFEOTK 163
Db 167 PCPSEPNTPTARTFDMVKRNPPKTKVSBPGLSPSGLTNTFTTQLTELEKEFHFNK 226
QY 164 YLAGPERARLAYSIGMTESQVWVFQNRRTKWRKSALEPSSSTPRAPGASGDRAASEN 223
Db 227 YLSRARRVEIAATLEINETQVKINFORNMKQKRE--REGGRVPPAPPGCPKEAAGDAS 284
QY 224 EDDEYNKP 231
Db 285 DQSTCTSP 292

RESULT 36
US-09-095-117-8
; Sequence 8, Application US/09095117
; Patent No. 6228582
; GENERAL INFORMATION:
; APPLICANT: Rodier, Patricia M.
; APPLICANT: Ingram, Jennifer L.
; APPLICANT: Figlewicz, Denise A.
; APPLICANT: Hyman, Susan L.
; APPLICANT: Stodgell, Christopher J.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS WHICH ARE
; TITLE OF INVENTION: ASSOCIATED WITH AUTISM SPECTRUM DISORDERS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:

Qy 146 TTTGHIQIFALEKTEQTKYLGPERRARLAYSLGWTESQVWVFQNRRTKWRKKSALPSS 205
Db 118 VFTELQMLGKRFKQKYLSTPDRIIDLAESLGLSQLQVKTWYQNRMRKWKIVLQGGGL 177
Qy 206 STPRAPGG-----ASGDRAASENEDDEYNKPLD-PDSDDKIR 242
Db 178 ESPTKPKGRPKKNSIPTSEQLTEQERAKDAEPAEVPGEPSDRSR 222

RESULT 39
US-09-949-016-7007
; Sequence 7007, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7007
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7007

Query Match 13.0%; Score 181.5; DB 4; Length 225;
Best Local Similarity 28.4%; Pred. No. 6.3e-11;
Matches 64; Conservative 29; Mismatches 97; Indels 35; Gaps 10;
Qy 40 PPGGLPQLAAGT-----PHGITDIL-SRP-----VATPNSLSLSGYPHV-AGFGGLSSQ 86
Db 12 PKGAAPAAAAGAGELLKFGVQALLAARPFHSHLAVLKAEOAVFKFPLAPLGCGLSSA 71
Qy 87 GVIYGPVGSFSGKAGNEY-PTRTRNCWADTQDWRGSRAPCGNTPDPLSDTIHKKHTRP 145
Db 72 LLAAGP--GLPGAGAPHLPLELQ-----LKGLEAAG--PGEFGTKAKKGRSRT 118
Qy 146 TTTGHIQIFALEKTEQTKYLGPERRARLAYSLGWTESQVWVFQNRRTKWRKKSALPSS 205
Db 119 VFTELQMLGKRFKQKYLSTPDRIIDLAESLGLSQLQVKTWYQNRMRKWKIVLQGGGL 178
Qy 206 STPRAPGG-----ASGDRAASENEDDEYNKPLD-PDSDDKIR 242
Db 179 ESPTKPKGRPKKNSIPTSEQLTEQERAKDAEPAEVPGEPSDRSR 223

RESULT 40
US-09-031-962D-2
; Sequence 2, Application US/09031962D
; Patent No. 6350867
; GENERAL INFORMATION:
; APPLICANT: Thomas C. Hart
; APPLICANT: Jennifer A. Price
; TITLE OF INVENTION: Methods and Compositions for Enhancing
; TITLE OF INVENTION: Osseous Growth, Repair, and Regeneration
; FILE REFERENCE: WFU98-18
; CURRENT APPLICATION NUMBER: US/09/031,962D
; CURRENT FILING DATE: 1998-02-27
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-031-962D-2
Query Match 12.9%; Score 180.5; DB 3; Length 287;
Best Local Similarity 26.2%; Pred. No. 1.2e-10;
Matches 56; Conservative 32; Mismatches 55; Indels 71; Gaps 9;
Qy 14 TQLAQFSEMKAPMCQYVSQNSFYKLSPPGLGPQLAAGTPHGITTILSRPVATPNSLSLSG 73
Db 38 TDLGYYS---AP-----QHDYY-----SQPYGQT-----VNPYT----- 64
Qy 74 YPHVAGFGGLSSQGVVYYPQVGSFSGKAGNEYPTTRNCWADTQDWRGSRAPCGNTPDPL 133
Db 65 YHOFNLNGLAGTCAYSKSEYTYGASYRQY-----GAYR---EQPLPA 105
Qy 134 SDTI-----HKKKHTRPTTTGHIQIFALEKTEQTKYLGPERRARLAY 176
Db 106 QDPVSVKEEPEAEVRMVGPKKVRKPTTIYSSVQLAALQRRFQKQYALPERAELAAQ 165
Qy 177 LGMTESQVWVFQNRRTKWRK--KSALEPSSSTP 208
Db 166 LGLTQTVKIMFQNRRSKFKLYKNGEVPLEHSP 199
Search completed: April 13, 2005, 17:18:51
Job time : 71.1554 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: April 13, 2005, 17:02:44 ; Search time 4.89527 Seconds
(without alignments)
452.066 Million cell updates/sec

Title: US-09-998-861-7
Perfect score: 102
Sequence: 1 XTKTPFSVKDILSLPEQXRAXGA 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR-79:*
1: Pirl:*
2: Pirl2:*
3: Pirl3:*
4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60	58.8	269	2 I50504	nk2.2 protein - zebra fish
2	60	58.8	273	2 I48187	gene NKX2.2 protein - golden hamster
3	60	58.8	273	2 JC4634	vertebrate NK-2 homeobox protein -
4	56	54.9	299	2 I51442	homeobox protein -
5	55	53.9	416	2 A43561	homeotic protein m
6	46.5	45.6	371	2 I46089	thyroid transcript
7	46.5	45.6	372	2 S53724	thyroid-specific
8	46.5	45.6	372	2 S12002	thyroid nuclear fa
9	46.5	45.6	401	2 G02321	thyroid transcript
10	46	45.1	364	2 I48188	gene NKX6.1 protei
11	44	43.1	495	2 B70322	hypothetical prote
12	44	43.1	553	2 AH0849	invasion protein r
13	44	43.1	553	2 S70817	invasion genes tra
14	44	43.1	692	2 T41240	hypothetical zinc
15	44	43.1	2287	2 T21312	hypothetical prote
16	43.5	42.6	551	2 D96724	hypothetical prote
17	43	42.2	725	2 H83586	malate synthase G
18	43	42.2	991	2 S57385	probable membrane
19	43	42.2	1888	2 T39009	hypothetical prote
20	42	41.2	61	2 PS0407	muscle segment hom
21	42	41.2	128	2 S20896	homeotic protein m
22	42	41.2	155	2 T20024	hypothetical prote
23	42	41.2	197	2 T31286	2-nitrotoluene dio
24	42	41.2	219	2 D70333	hypothetical prote
25	42	41.2	453	2 AB7431	alpha-glucosidase-b
26	42	41.2	453	2 AB2649	hypothetical prote
27	42	41.2	481	2 D71927	cag island protein
28	42	41.2	593	2 P75032	rnase l inhibitor
29	42	41.2	615	2 F87599	hypothetical prote

30	42	41.2	798	2 B27079	fibronectin recept
31	42	41.2	1873	2 T30944	surface protein pr
32	42	41.2	6359	2 T31679	bacitracin synthet
33	41	40.2	61	2 PS0404	muscle segment hom
34	41	40.2	79	2 G34510	homeotic protein H
35	41	40.2	105	2 E81956	PEMK-like protein
36	41	40.2	107	2 C81014	PEMK-related prote
37	41	40.2	335	2 AH0900	probable monooxyge
38	41	40.2	363	2 C86214	hypothetical prote
39	41	40.2	431	2 AH2982	metallo-beta-lacta
40	41	40.2	431	2 G98300	hypothetical prote
41	41	40.2	437	2 S55392	msh protein - fruit
42	41	40.2	590	2 G71136	probable transport
43	41	40.2	722	2 H97217	uncharacterized co
44	41	40.2	1192	2 T17255	hypothetical prote
45	41	40.2	1210	2 D88013	protein K10B4.1 [i

ALIGNMENTS

RESULT 1
I50504

nk2.2 protein - zebra fish
C:Species: Brachydanio rerio (zebra fish)
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 16-Aug-2004
C:Accession: I50504
R:Barth, K.A.; Wilson, S.W.
Development 121, 1755-1768, 1995
A>Title: Expression of zebrafish nk2.2 is influenced by sonic hedgehog/vertebrate hedgehog
A:Reference number: I50504; MUID:95324401; PMID:7600991
A:Accession: I50504
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-269 <BAR>
A:Cross-references: UNIPROT:Q90481; EMBL:X85977; NID:g999448; PIDN:CAAS59967.1; PID:g9999.
C:Genetics:
A:Gene: nk2.2
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:126-182/Domain: homeobox homology <HOX>

Query Match 58.8%; Score 60; DB 2; Length 269;
Best Local Similarity 59.1%; Pred. No. 0.02; Mismatches 2; Indels 0; Gaps 0;
Matches 13; Conservative 7; Indels 0; Gaps 0;

QY 2 TKTTPFSVKDILSLPEQXRAXGA 23
||| ||||| |||
DB 6 TKTGFSVKDILDLPTNDEGS 27

RESULT 2
I48187

gene NKX2.2 protein - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Aug-2004
C:Accession: I48187
R:Rudnick, A.; Ling, T.Y.; Odagiri, H.; Rutter, W.J.; German, M.S.
Proc. Natl. Acad. Sci. U.S.A. 91, 12203-12207, 1994
A>Title: Pancreatic beta cells express a diverse set of homeobox genes.
A:Reference number: I48185; MUID:95083670; PMID:7991607
A:Accession: I48187
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-273 <RES>
A:Cross-references: UNIPROT:P43697; EMBL:X81408; NID:g587464; PIDN:CAAS7165.1; PID:g587.
C:Genetics:
A:Gene: NKX2.2
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:129-185/Domain: homeobox homology <HOX>

Query Match 58.8%; Score 60; DB 2; Length 273;

Best Local Similarity 59.1%; Pred. No. 0.02;
Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 TKTPTSVKDIILSLPEQXRAXGA 23
||| ||||| |||
DB 6 TKTGFSVKDILDLPTNDEGGS 27

RESULT 3

JC4634
vertebrate NK-2 homeobox protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Apr-1996 #sequence revision 24-May-1996 #text_change 16-Aug-2004
C:Accession: JC4634; I49349; PH0860
R:Hartigian, D.J.; Rubenstein, J.L.R.
Gene 168, 271-272, 1996
A:Title: The cDNA sequence of murine Nkx-2.2.
A:Reference number: JC4634; MUID:96194911; PMID:8654958
A:Accession: JC4634
A:Molecule type: mRNA
A:Residues: 1-273 <HAR>
A:Cross-references: UNIPROT:P42586; GB:U31566; NID:G1019863; PIDN:AAA79303.1; PID:G1019863
A:Experimental source: embryonic telencephalon
R:Rubenstein, J.L.; Martinez, S.; Shimamura, K.; Puelles, L.
Science 266, 578-580, 1994
A:Title: The embryonic vertebrate forebrain: the prosomeric model.
A:Reference number: I49349; MUID:95025968; PMID:7939711
A:Accession: I49349
A:Status: preliminary; nucleic acid sequence not shown; translation not shown; translated
A:Molecule type: mRNA
A:Residues: 1-273 <RES>
A:Cross-references: EMBL:U31566; NID:G1019863; PIDN:AAA79303.1; PID:G1019864
R:Price, M.; Lazzaro, D.; Pohl, T.; Mattei, M.G.; Ruether, U.; Olivo, J.C.; Duboule, D.;
Neuron 8, 241-255, 1992
A:Title: Regional expression of the homeobox gene Nkx-2.2 in the developing mammalian fo
A:Reference number: PH0860; MUID:92153416; PMID:1346742
A:Accession: PH0860
A:Molecule type: DNA
A:Residues: 128-187; 199-223 <PRI>
A:Experimental source: brain
C:Genetics:
A:Gene: Nkx-2.2
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:129-185/Domain: homeobox homology <HGX>

Query Match 58.8%; Score 60; DB 2; Length 273;
Best Local Similarity 59.1%; Pred. No. 0.02;
Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 TKTPTSVKDIILSLPEQXRAXGA 23
||| ||||| |||
DB 6 TKTGFSVKDILDLPTNDEGGS 27

RESULT 4

I51442
homeobox protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 13-Sep-1996 #sequence revision 13-Sep-1996 #text_change 16-Aug-2004
C:Accession: I51442
R:Tonissen, K.F.; Drysdale, T.A.; Lints, T.J.; Krieg, P.A.
Dev. Biol. 162, 325-328, 1994
A:Title: Nkx-2.5: A Xenopus gene related to Nkx-2.5 and tinman: Evidence for a conserve
A:Reference number: I51441; MUID:94170922; PMID:7545912
A:Accession: I51442
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-299 <TON>
A:Cross-references: UNIPROT:P42583; GB:L25600; NID:G409372; PIDN:AAA19861.1; PID:G409372
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:125-181/Domain: homeobox homology <HGX>

Query Match 54.9%; Score 56; DB 2; Length 299;
Best Local Similarity 73.3%; Pred. No. 0.11; Mismatches 2; Indels 0; Gaps 0;

QY 2 TKTPTSVKDIILSLPE 16
||| ||||| |||
DB 7 TSTPTSVKDIILNLEQ 21

RESULT 5

A43561
homeotic protein msh-2 - fruit fly (Drosophila melanogaster)
N:Alternate names: homeotic protein NK-4
C:Species: Drosophila melanogaster
C:Date: 11-Dec-1992 #sequence revision 11-Dec-1992 #text_change 09-Jul-2004
C:Accession: A43561; D33976
R:Bodmer, R.; Jan, L.Y.; Jan, Y.N.
Development 110, 661-669, 1990
A:Title: A new homeobox-containing gene, msh-2, is transiently expressed early during me
A:Reference number: A43561; MUID:91209226; PMID:1982429
A:Accession: A43561
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-416 <BOD>
A:Cross-references: UNIPROT:P22711; GB:X55192; NID:G10345; PIDN:CAA38978.1; PID:G10346
R:Kim, Y.; Nirenberg, M.
Proc. Natl. Acad. Sci. U.S.A. 86, 7716-7720, 1989
A:Title: Drosophila NK-homeobox genes.
A:Reference number: A33976; MUID:90046666; PMID:2573058
A:Accession: D33976
A:Molecule type: DNA
A:Residues: 267-416 <KIM>
A:Cross-references: GB:M27292; NID:G157639; PIDN:AAA28619.1; PID:G552094
C:Genetics:
A:Gene: FlyBase:tin
A:Cross-references: FlyBase:FBgn0004110
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:302-358/Domain: homeobox homology <HGX>

Query Match 53.9%; Score 55; DB 2; Length 416;
Best Local Similarity 64.7%; Pred. No. 0.23;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 TPTSVKDIILSLPEQXRA 20
||| ||||| |||
DB 36 TPTSVKDIILNVMNTEA 52

RESULT 6

I46089
thyroid transcription factor-1 - dog
C:Species: Canis lupus familiaris (dog)
C:Date: 16-Aug-1996 #sequence revision 16-Aug-1996 #text_change 16-Aug-2004
C:Accession: I46089; S42104
R:Van Renterghem, P.H.G.; Dremier, S.; Vassar, G.; Christophe, J.
Mol. Cell. Endocrinol. 112, 83-93, 1995
A:Title: Study of TTF1 gene expression in dog thyrocytes in primary culture.
A:Reference number: I46089; MUID:96034516; PMID:7589789
A:Accession: I46089
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Residues: 1-371 <VAN>
A:Cross-references: UNIPROT:P43698; EMBL:X77910; NID:G457488; PIDN:CAAS4868.1; PID:G457488
C:Genetics:
A:Gene: TTF-1
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:162-218/Domain: homeobox homology <HGX>

Query Match 45.6%; Score 46.5; DB 2; Length 371;
Best Local Similarity 60.0%; Pred. No. 5.6;

Matches 12; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 4 TPFSVKDILS-LPEQXRAXG 22
||||| ||||| :
DB 9 TPFSVSDILSPLSEYKKVG 28

RESULT 7

thyroid-specific enhancer-binding protein (T/EBP) - mouse
C:Species: Mus musculus (house mouse)
C>Date: 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 16-Aug-2004
C:Accession: S53724
R:Oguchi, H.; Pan, Y.T.; Kimura, S.
Biochim. Biophys. Acta 1261, 304-306, 1995
A:Title: The complete nucleotide sequence of the mouse thyroid-specific enhancer-binding
A:Reference number: S53724; MUID:95226463; PMID:7711079
A:Accession: S53724
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-372 <OGU>
A:Cross-references: UNIPROT:P50220; GB:U19755; NID:9885887; PIDN:AAA86100.1; PID:9885888
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
P:162-218/Domain: homeobox homology <HOX>

Query Match 45.6%; Score 46.5; DB 2; Length 372;
Best Local Similarity 60.0%; Pred. No. 5.6;

Matches 12; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 4 TPFSVKDILS-LPEQXRAXG 22
||||| ||||| :
DB 9 TPFSVSDILSPLSEYKKVG 28

RESULT 8

thyroid nuclear factor 1 - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 13-Jan-1995 #sequence_revision 13-Mar-1997 #text_change 16-Aug-2004
C:Accession: S12002; PC2252
R:Guazzi, S.; Price, M.; de Felice, M.; Damante, G.; Mattei, M.G.; di Lauro, R.
EMBO J. 9, 3631-3639, 1990
A:Title: Thyroid nuclear factor 1 (TFP-1) contains a homeodomain and displays a novel DN
A:Reference number: S12002; MUID:91006063; PMID:1976511
A:Accession: S12002
A:Molecule type: mRNA
A:Residues: 1-372 <GU>
A:Cross-references: UNIPROT:P23441; EMBL:X53858; NID:957422; PIDN:CAA37851.1; PID:957423
A>Note: Met-1 was determined by sequencing DNA
R:Endo, T.; Ohta, K.; Saito, T.; Haraguchi, K.; Nakazato, M.; Kogai, T.; Onaya, T.
Biochem. Biophys. Res. Commun. 204, 1358-1363, 1994
A:Title: Structure of the rat thyroid transcription.
A:Reference number: PC2252; MUID:95071477; PMID:7980615
A:Accession: PC2252
A:Molecule type: DNA
A:Residues: 1-35 <END>
A:Cross-references: DBJ:D38035
A:Experimental source: liver
C:Comment: This protein activates the genes of thyroid specific protein, thyroglobulin,
C:Genetics:
A:Gene: TTF-1
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
P:162-218/Domain: homeobox homology <HOX>

Query Match 45.6%; Score 46.5; DB 2; Length 372;
Best Local Similarity 60.0%; Pred. No. 5.6;

Matches 12; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 4 TPFSVKDILS-LPEQXRAXG 22
||||| ||||| :
DB 9 TPFSVSDILSPLSEYKKVG 28

RESULT 9

thyroid transcription factor 1 - human
N:Alternate names: Drosophila NK-2 homolog A; thyroid nuclear factor
C:Species: Homo sapiens (man)
C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 16-Aug-2004
C:Accession: G02321; G02039; A56451; S53723; S53725; G02041
R:Hamdan, H.
submitted to the EMBL Data Library, December 1995
A:Reference number: H01047
A:Accession: G02321
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-401 <HAN>
A:Cross-references: UNIPROT:Q9BRJ8; EMBL:U43203; NID:g1199864; PIDN:AAA89066.1; PID:g1199864
R:Hamdan, H.
submitted to the EMBL Data Library, August 1995
A:Reference number: G09098
A:Accession: G02039
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-71 <HA>
A:Cross-references: EMBL:U33627; NID:g1113816; PID:g1000129
R:Ikeda, K.; Clark, J.C.; Shaw-White, J.R.; Stahlman, M.T.; Boutell, C.J.; Whitsett, J.
J. Biol. Chem. 270, 8108-8114, 1995
A:Title: Gene structure and expression of human thyroid transcription factor-1 in respi,
A:Reference number: A56451; MUID:95229626; PMID:7713914
A:Accession: A56451
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 31-401 <IKE>
A:Cross-references: GB:U19816; NID:g767832; PIDN:AAC50125.1; PID:g767833
R:Saiardi, A.; Taasi, V.; de Filippis, V.; Civitareale, D.
Biochim. Biophys. Acta 1261, 307-310, 1995
A:Title: Cloning and sequence analysis of human thyroid transcription factor 1.
A:Reference number: S53723; MUID:95226464; PMID:7711080
A:Accession: S53723
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 31-401 <SAI>
A:Cross-references: EMBL:X82850; NID:9695582; PIDN:CAA58053.1; PID:9695583
R:Oguchi, H.; Pan, Y.T.; Kimura, S.
Biochim. Biophys. Acta 1261, 304-306, 1995
A:Title: The complete nucleotide sequence of the mouse thyroid-specific enhancer-binding
A:Reference number: S53724; MUID:95226463; PMID:7711079
A:Accession: S53725
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 31-271, 'G', 272-401 <OGU>
R:Hamdan, H.
submitted to the EMBL Data Library, August 1995
A:Reference number: G09102
A:Accession: G02041
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 31-255, 'HB', 258-401 <HA3>
A:Cross-references: EMBL:U33749; NID:g1072047; PID:g1072048
C:Genetics:
A:Gene: TTF-1; GDB:TTF1; NKX2A
A:Cross-references: GDB:132588; OMIM:600635
A:Map position: 14q13-14q13
A:Introns: 26/2; 155/1
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
P:192-248/Domain: homeobox homology <HOX>

Query Match 45.6%; Score 46.5; DB 2; Length 401;
Best Local Similarity 60.0%; Pred. No. 5.1;

Matches 12; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 4 TPFSVKDILS-LPEQXRAXG 22

DB 39 TPFSVDILSPLEESYKKVG 58
||||| ||||| ||||| ||||| : |

RESULT 10

T48188
gene NKx6.1 protein - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Aug-2004
C:Accession: I48188
R:Rudnick, A.; Ling, T.Y.; Odagiri, H.; Rutter, W.J.; German, M.S.
Proc. Natl. Acad. Sci. U.S.A. 91, 12203-12207, 1994
A>Title: Pancreatic beta cells express a diverse set of homeobox genes.
A:Reference number: I48185; MUID:95083670; PMID:7991607
A:Accession: I48188
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-364 <RES>
A:Cross-references: UNIPROT:Q60554; EMBL:X81409; NID:9587466; PIDN:CAA57166.1; PID:95874
C:Genetics:
A:Gene: NKx6.1
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:237-293/Domain: homeobox homology <HOX>

Query Match 45.1%; Score 46; DB 2; Length 364;
Best Local Similarity 50.0%; Pred. No. 6.7;
Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 4 TPFSVKDILSPLEQXAXGA 23
|| : ||||| |||||

DB 93 TPFGINDILSRPMPVAGSA 112
|| : ||||| |||||

RESULT 11

B70322
Hypothetical protein aq_243 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C:Accession: B70322
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O.V.
Nature 392, 353-358, 1998
A>Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: B70322
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-495 <AQF>
A:Cross-references: UNIPROT:Q66606; GB:AE000680; NID:92982948; PIDN:AAC06570.1; PID:9298
A:Experimental source: strain VFS
C:Genetics:
A:Gene: aq_243
C:Superfamily: Aquifex aeolicus hypothetical protein aq_243

Query Match 43.1%; Score 44; DB 2; Length 495;
Best Local Similarity 60.0%; Pred. No. 21;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 9 KDILSLPEQXAXGA 23
|| : ||||| |||||

DB 326 KDALNLPEDFRVDGA 340
|| : ||||| |||||

RESULT 12

AH0849
Invasion protein regulator iagA [imported] - Salmonella enterica subsp. enterica serovar
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A>Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AH0849
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AH0849
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-553 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05983.1; PID:g16503954; GSPDB:GN00176
C:Genetics:
A:Gene: iagA

Query Match 43.1%; Score 44; DB 2; Length 553;
Best Local Similarity 61.1%; Pred. No. 23;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 TKTPFSVKDILSLPEQXR 19
|| : ||||| |||||

DB 164 TKNCRSVKDILELMDQLR 181
|| : ||||| |||||

RESULT 13

S70817
invasion genes transcription activator hila - Salmonella typhimurium
C:Species: Salmonella typhimurium
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C:Accession: S70817; S70816
R:Bajaj, V.; Hwang, C.; Lee, C.A.
submitted to the EMBL Data Library, April 1995
A:Description: hila is a novel ompR/coxR family member that activates the expression of
A:Reference number: S70817
A:Accession: S70817
A:Molecule type: DNA
A:Residues: 1-553 <BAJ>
A:Cross-references: UNIPROT:P43015; EMBL:U25352; NID:g1050873; PIDN:AAD12579.1; PID:g806
A:Experimental source: strain SL1344
A>Note: it is uncertain whether Met-1 or Met-23 is the initiator
R:Bajaj, V.; Hwang, C.; Lee, C.A.
Mol. Microbiol. 18, 715-727, 1995
A>Title: hila is a novel ompR/coxR family member that activates the expression of Salmon
A:Reference number: S70816; MUID:96414472; PMID:8817493
A:Accession: S70816
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 23-27,29-110 <BA2>
A:Cross-references: EMBL:U25352
A:Experimental source: strain SL1344
C:Genetics:
A:Gene: hila
C:Keywords: DNA binding; transcription regulation

Query Match 43.1%; Score 44; DB 2; Length 553;
Best Local Similarity 61.1%; Pred. No. 23;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 TKTPFSVKDILSLPEQXR 19
|| : ||||| |||||

DB 164 TKNCRSVKDILELMDQLR 181
|| : ||||| |||||

RESULT 14

T41240
hypothetical zinc finger protein SPC1919.15 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T41240; T41606
R:Lyne, M.; Harris, D.E.; Murphey, L.D.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 1999
A:Reference number: Z21979
A:Accession: T41240
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-692 <LYN1>
A:Cross-references: UNIPROT:O74924; EMBL:AL035075; NID:g4107303; PIDN:CAA22646.1; PID:g4
A:Experimental source: strain 972h-; cosmid c1919
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Hilbert, H.; Duesterhoeft, A.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z22004
A:Accession: T41606
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 594-692 <LYN2>
A:Cross-references: EMBL:AL031855; NID:g3738198; PIDN:CAA21291.1; PID:g3738199; GSPDB:GN
A:Experimental source: strain 972h-; cosmid c790
C:Genetics:
A:Gene: SPDB:SPCC1919.15; SPDB:SPCC790.01
A:Map position: 3

Query Match 43.1%; Score 44; DB 2; Length 692;
Best Local Similarity 66.7%; Pred. No. 30;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 KTFPSVKDILSL 14
:|||||:|:
Db 679 ETPFSVSDILTI 690

RESULT 15
T21312
hypothetical protein F23D12.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T21312; T24907
R:Barlow, K.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19404
A:Accession: T21312
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2287 <W1>
A:Cross-references: UNIPROT:Q19761; EMBL:Z71186; PIDN:CAA94917.1; GSPDB:GN000028; CESP:F2
A:Experimental source: clone F23D12
R:Barlow, K.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19952
A:Accession: T24907
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2287 <W12>
A:Cross-references: EMBL:Z70687; NID:g1256502; PIDN:CAA94618.1; GSPDB:GN000028; CESP:F23D
A:Experimental source: clone T14C1
C:Genetics:
A:Gene: CESP:F23D12.2
A:Map position: X
A:Introns: 75/3; 141/3; 165/3; 409/3; 497/3; 572/1; 583/3; 686/3; 773/3; 916/3; 945/3; 1

Query Match 43.1%; Score 44; DB 2; Length 2287;
Best Local Similarity 69.2%; Pred. No. 11e+02;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 KTFPSVKDILSLP 15
:|||||:|:
Db 265 KTFPSVNGVSLP 277

RESULT 16
D96724
hypothetical protein F20P5.12 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: D96724
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.P.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96724
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-551 <STO>
A:Cross-references: UNIPROT:O04530; GB:AE005173; NID:g2194125; PIDN:AA61100.1; GSPDB:G
C:Genetics:
A:Gene: F20P5.12
A:Map position: 1
C:Superfamily: Arabidopsis thaliana hypothetical protein F10M23.360

Query Match 42.6%; Score 43.5; DB 2; Length 551;
Best Local Similarity 62.5%; Pred. No. 28;
Matches 10; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 3 KTFPSVKDIL-SLPEQ 17
:|||||:|:
Db 26 KVPFSVNDVLPMLPRQ 41

RESULT 17
H83586
malate synthase G PA0482 [imported] - Pseudomonas aeruginosa (strain PAO1)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: H83586
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic path
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H83586
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-725 <STO>
A:Cross-references: UNIPROT:Q91636; GB:AE004485; GB:AE004091; NID:g9946332; PIDN:AA038,
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: glcB; PA0482

Query Match 42.2%; Score 43; DB 2; Length 725;
Best Local Similarity 61.5%; Pred. No. 46;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 KTFPSVKDILSLP 15
:|||||:|:
Db 568 RTPASVDLITIP 580

RESULT 18
S57385
probable membrane protein YOL084w - Yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein O0953
C:Species: Saccharomyces cerevisiae
C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jul-2004
C:Accession: S57385; S66778; S50421
R:Zumstein, E.; Pearson, B.M.; Kalogeropoulos, A.; Schweizer, M.
Yeast 11, 975-986, 1995
A:Title: A 29.425 kb segment on the left arm of yeast chromosome XV contains more than r
A:Reference number: S57374; MUID:96021609; PMID:8533473
A:Accession: S57385
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-991 <ZUM>
A:Cross-references: UNIPROT:Q12252; EMBL:X83121; NID:g600461; PIDN:CAA58195.1; PID:g600
R:Zumstein, E.; Pearson, B.M.; Kalogeropoulos, A.; Schweizer, M.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S66775
A:Accession: S66778
A:Molecule type: DNA
A:Residues: 1-991 <ZOW>
A:Cross-references: EMBL:Z74826; NID:g1419921; PIDN:CAA990996.1; PID:g1419922; MIPS:YOL08
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:PHW7
A:Cross-references: SGD:S0005444
A:Map position: 15L
C:Superfamily: yeast probable membrane protein YOL084w
C:Keywords: transmembrane protein
F:14-30/Domain: transmembrane #status predicted <TM1>
F:95-111/Domain: transmembrane #status predicted <TM2>
F:141-157/Domain: transmembrane #status predicted <TM3>
F:392-408/Domain: transmembrane #status predicted <TM4>
F:442-458/Domain: transmembrane #status predicted <TM5>
F:480-496/Domain: transmembrane #status predicted <TM6>
F:585-601/Domain: transmembrane #status predicted <TM7>
F:641-657/Domain: transmembrane #status predicted <TM8>
F:665-681/Domain: transmembrane #status predicted <TM9>

Query Match 42.2%; Score 43; DB 2; Length 991;
Best Local Similarity 53.3%; Pred. NO. 65;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 5 PPSVKDILSLPEQXR 19

Db 41 PRSLXDIQTIPPEER 55

RESULT 19

T39009
hypothetical protein SPAC6B12.02c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
A:Accession: T39009
R:Gentles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1995
A:Reference number: Z21815
A:Accession: T39009
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1888 <GEN>
A:Cross-references: UNIPROT:O14207; EMBL:Z98531; PIDN:CAB11064.1; GSPDB:GN00066; SPDB:SB
A:Experimental source: strain 972h-; cosmid c6B12
C:Genetics:
A:Gene: SPDB:SPAC6B12.02c
A:Map position: 1
C:Superfamily: Schizosaccharomyces hypothetical protein SPAC6B12.02c

Query Match 42.2%; Score 43; DB 2; Length 1888;
Best Local Similarity 58.3%; Pred. NO. 13e+02;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 KTFPSVKDILSL 14

Db 1084 ESPFSIKDFMSL 1095

RESULT 20

PS0407
muscle segment homeotic protein B - zebra fish (fragment)
C:Species: Brachydanio rerio (zebra fish)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Aug-2004
A:Accession: PS0407
R:Holland, P.W.H.
Gene 98, 253-257, 1991
A:Title: Cloning and evolutionary analysis of msh-like homeobox genes from mouse, zebrafish
A:Reference number: PS0404; MUID:91200674; PMID:1673109
A:Accession: PS0407
A:Molecule type: DNA

A:Residues: 1-61 <HOL>
A:Cross-references: UNIPROT:Q03356; GB:M38579; NID:g214989; PIDN:AAA50032.1; PID:g214990
C:Genetics:
A:Gene: msh
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:3-59/Domain: homeobox homology <HOX>

Query Match 41.2%; Score 42; DB 2; Length 61;
Best Local Similarity 47.1%; Pred. NO. 4.5;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 KTFPSVKDILSLPEQXR 19

Db 6 RTPFSTQSLLSLERKFR 22

RESULT 21

S20896
homeotic protein msh - Chlorohydra viridissima (fragment)
C:Species: Chlorohydra viridissima
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-2004
A:Accession: S20896; S19822
R:Schummer, M.; Scheurien, I.; Schaller, C.; Galliot, B.
EMBO J. 11, 1815-1823, 1992
A:Title: HOX/HOX homeobox genes are present in hydra (Chlorohydra viridissima) and are
A:Reference number: S20892; MUID:92258391; PMID:1374713
A:Accession: S20896
A:Molecule type: mRNA
A:Residues: 1-128 <SCH>
A:Cross-references: UNIPROT:Q23824; EMBL:X64629; NID:g7129; PIDN:CAA45912.1; PID:g7130
C:Genetics:
A:Gene: msh
C:Superfamily: homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:18-74/Domain: homeobox homology <HOX>

Query Match 41.2%; Score 42; DB 2; Length 128;
Best Local Similarity 41.2%; Pred. NO. 10;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 3 KTFPSVKDILSLPEQXR 19

Db 21 RTPFSVNQLLTLEQKFK 37

RESULT 22

T20024
hypothetical protein C47F8.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
A:Accession: T20024
R:McLay, K.
submitted to the EMBL Data Library, December 1997
A:Reference number: Z19212
A:Accession: T20024
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-155 <WIL>
A:Cross-references: UNIPROT:O62110; EMBL:AL009246; PIDN:CAA15834.1; GSPDB:GN00019; CESP
A:Experimental source: Clone C47F8
C:Genetics:
A:Gene: CESP:C47F8.7
A:Map position: 1
A:Introns: 52/1; 83/1
C:Superfamily: Caenorhabditis elegans hypothetical protein C47F8.7

Query Match 41.2%; Score 42; DB 2; Length 155;
Best Local Similarity 50.0%; Pred. NO. 13;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 TKTFPSVKDILSLPEQ 17

:::|||||:::|||||

A:Accession: A97431
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-453 <KUR>
A:Cross-references: UNIPROT:Q8UHT8; GB:AE007869; PIDN:AAK86402.1; PID:gl15155534; GSPDB:
C:Genetics:
A:Gene: AGR_C1045
A:Map position: circular chromosome

Query Match 41.2%; Score 42; DB 2; Length 453;
Best Local Similarity 45.0%; Pred. No. 41;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 3 KTFPSVKDILSLPEQXRAGX 22
Db 187 KVPESMEDLLKLTQDIVADG 206

RESULT 26
AB2649
hypothetical protein agle [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AB2649
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayin, T.; Levy, R.; Li, M.; McClellan,
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Kraspan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AB2649
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-453 <KUR>
A:Cross-references: UNIPROT:Q8UHT8; GB:AE008688; PIDN:AAL41608.1; PID:gl17738945; GSPDB:
C:Genetics:
A:Gene: agle
A:Map position: circular chromosome

Query Match 41.2%; Score 42; DB 2; Length 453;
Best Local Similarity 45.0%; Pred. No. 41;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 3 KTFPSVKDILSLPEQXRAGX 22
Db 187 KVPESMEDLLKLTQDIVADG 206

RESULT 27
D71927
cag island protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C:Accession: D71927
R:Am, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric patho
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: D71927
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-481 <ARN>
A:Cross-references: UNIPROT:Q9ZLV5; GB:AE001481; GB:AE001439; NID:g4155005; PIDN:AAD060
C:Genetics:
A:Gene: orf8
A:Map position: circular chromosome

Query Match 41.2%; Score 42; DB 2; Length 481;


```

Db      13 RTPTTQQLLSLEKKFR 29
||||: : |||| :|
|

RESULT 35
E81956
PEM-K-like protein NMA0400 [imported] - Neisseria meningitidis (strain Z2491 serogroup A)
C/Species: Neisseria meningitidis
C/Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 09-Jul-2004
C/Accession: E81956
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel,
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A/Reference number: A81775; MUID:20222556; PMID:10761919
A/Accession: E81956
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-105 <PAR>
A/Cross-references: UNIPROT:Q9JWF2; GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB8370
A/Experimental source: serogroup A, strain Z2491
C/Genetics:
A/Gene: chpA; NMA0400

Query Match 40.2%; Score 41; DB 2; Length 105;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 FSVKDIILSLPEQXRA 20
| | | | | | | |
| | | | | | | |

Db 61 FQDKDGLLLPEQIRA 75
| | | | | | | |
| | | | | | | |

RESULT 36
C81014
PEM-k-related protein NMB2038 [imported] - Neisseria meningitidis (strain MC58 serogroup
C/Species: Neisseria meningitidis
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C/Accession: C81014
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.;
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizzi, M.
Science 287, 1809-1815, 2000
A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A/Reference number: A81000; MUID:20175755; PMID:10710307
A/Accession: C81014
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-107 <TET>
A/Cross-references: UNIPROT:Q9JX11; GB:AE002553; GB:AE002098; NID:g7227292; PIDN:AAF423
C/Genetics:
A/Gene: NMB2038

Query Match 40.2%; Score 41; DB 2; Length 107;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 FSVKDIILSLPEQXRA 20
| | | | | | | |
| | | | | | | |

Db 63 FQDKDGLLLPEQIRA 77
| | | | | | | |
| | | | | | | |

RESULT 37
AH0900
probable monooxygenase [imported] - Salmonella enterica subsp. enterica serovar Typhi (t
C/Species: Salmonella enterica subsp. enterica serovar Typhi
A/Note: this species has also been called Salmonella typhi
C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C/Accession: AH0900
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

```


S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
 A:Reference number: AB0502; MUID:21534947; PMID:11677608
 A:Accession: AH0900
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-335 <PAR>
 A:CROSS-references: GB:AL513382; PIDN:CAD07798.1; PID:gl6504346; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY3459
 C:Superfamily: ynbw protein

Query Match 40.2%; Score 41; DB 2; Length 335;
 Best Local Similarity 47.4%; Pred. No. 43;
 Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 5 PFSVKDILSLPQXRXGXA 23
 ||||| | : ||| |
 Db 7 PFSVLDLAPIEGSSAKA 25

RESULT 38
 C86214
 hypothetical protein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
 A:Reference number: C86214
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: C86214
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-363 <STO>
 A:CROSS-references: UNIPROT:Q9LQO9; GB:AE005172; NID:g8439881; PIDN:AAF75067.1; GSPDB:GN
 C:Genetics:
 A:Map position: 1
 C:Superfamily: kinase-related transforming protein; protein kinase homology

Query Match 40.2%; Score 41; DB 2; Length 363;
 Best Local Similarity 58.3%; Pred. No. 47;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 8 VKDILSLPQXRX 19
 :|||: |||: |
 Db 95 IKDIIEPKEK 106

RESULT 39
 AH2982
 metallo-beta-lactamase superfamily protein [imported] - Agrobacterium tumefaciens (strain
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
 C:Accession: AH2982
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; MUID:21608550; PMID:11743193
 A:Accession: AH2982

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-431 <KUR>
 A:CROSS-references: UNIPROT:Q8UAA9; GB:AE008689; PIDN:AAL44278.1; PID:gl7741864; GSPDB:G
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu3465
 A:Map position: linear chromosome

Query Match 40.2%; Score 41; DB 2; Length 431;
 Best Local Similarity 53.8%; Pred. No. 57;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 7 SVKDILSLPQXRX 19
 |::|||: |||: |
 Db 328 SIQDILALPDDTR 340

RESULT 40
 G98300
 hypothetical protein AGR_L_2726 [imported] - Agrobacterium tumefaciens (strain C58, Cere
 C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
 C:Accession: G98300
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A:Reference number: A97359; MUID:21608551; PMID:11743194
 A:Accession: G98300
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-431 <KUR>
 A:CROSS-references: UNIPROT:Q8UAA9; GB:AE007870; PIDN:AAK89929.1; PID:gl5159882; GSPDB:G
 C:Genetics:
 A:Gene: AGR_L_2726
 A:Map position: linear chromosome

Query Match 40.2%; Score 41; DB 2; Length 431;
 Best Local Similarity 53.8%; Pred. No. 57;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 7 SVKDILSLPQXRX 19
 |::|||: |||: |
 Db 328 SIQDILALPDDTR 340

Search completed: April 13, 2005, 17:17:26
 Job time : 6.89527 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2005, 17:01:49 ; Search time 22.1453 Seconds
(without alignments)
531.844 Million cell updates/sec

Title: US-09-998-861-7

Perfect score: 102

Sequence: 1 XTKTPFSVKDILSLPEQXRAXCA 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_03.*

1: uniprot_sprot.*

2: uniprot_crembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	66.7	318	1 NK25 RAT	O35767 rattus norv
2	68	66.7	324	1 NK25 HUMAN	P52952 homo sapien
3	67	65.7	362	1 NK23 MOUSE	P97334 mus musculu
4	67	65.7	450	2 Q8BRG9	Q8brg9 mus musculu
5	65	63.7	318	1 NK25 MOUSE	P42582 mus musculu
6	65	63.7	318	1 NK25 MOUSE	Q925V3 mus musculu
7	65	63.7	321	2 Q98871	Q98871 brachydanio
8	62	60.8	335	2 Q91623	Q91623 xenopus lae
9	62	60.8	335	2 Q91861	Q91861 xenopus lae
10	61	59.8	329	2 Q91862	Q91862 xenopus lae
11	61	59.8	331	2 Q91863	Q91863 xenopus lae
12	61	59.8	364	1 NK23 HUMAN	Q8tau0 homo sapien
13	60	58.8	86	2 Q9HBZ5	Q9hbz5 homo sapien
14	60	58.8	86	2 Q9JXR0	Q9jkr0 mus musculu
15	60	58.8	143	2 Q6DGE7	Q6dge7 brachydanio
16	60	58.8	248	2 Q8BRG9	Q8brg9 mus musculu
17	60	58.8	269	1 NK22 BRARE	Q90481 brachydanio
18	60	58.8	269	2 Q98873	Q98873 brachydanio
19	60	58.8	269	2 Q6AXL1	Q6axl1 brachydanio
20	60	58.8	273	1 NK22 HUMAN	Q95096 homo sapien
21	60	58.8	273	1 NK22 MESAU	P43697 mesocricetu
22	60	58.8	273	1 NK22 MOUSE	P42586 mus musculu
23	60	58.8	294	1 NK25 CHICK	Q90788 gallus gall
24	59.5	58.3	193	2 Q13005	O13005 gallus gall
25	59.5	58.3	193	2 Q13053	O13053 gallus gall
26	58	56.9	54	2 Q6DN74	Q6dn74 xenopus lae
27	58	56.9	263	2 Q9W6H4	Q9w6h4 xenopus lae
28	57	55.9	312	2 Q96377	Q96377 brachioisto
29	57	55.9	394	2 Q6QB01	Q6qb01 strongyloce
30	56	54.9	77	2 Q9DDU7	Q9ddu7 xenopus lae
31	56	54.9	299	1 NK25_XENLA	P42583 xenopus lae

32 56 54.9 299 2 Q7TOT3 Q7tot3 xenopus lae
33 56 54.9 300 2 Q6DUE1 Q6due1 xenopus lae
34 55 53.9 252 2 Q6UKJ7 Q6ukj7 drosophila
35 55 53.9 252 2 Q6UKK8 Q6ukk8 drosophila
36 55 53.9 252 2 Q6UKL1 Q6ukl1 drosophila
37 55 53.9 252 2 Q6UKL4 Q6ukl4 drosophila
38 55 53.9 252 2 Q6UKL5 Q6ukl5 drosophila
39 55 53.9 252 2 Q6UKL7 Q6ukl7 drosophila
40 55 53.9 252 2 Q6UKL8 Q6ukl8 drosophila
41 55 53.9 416 1 HMTI_DROME P22711 drosophila
42 55 53.9 416 2 Q6AWK6 Q6awk6 drosophila
43 55 53.9 416 2 Q9VDA7 Q9vda7 drosophila
44 53 52.0 314 2 O13160 O13160 brachydanio
45 53 52.0 314 2 Q98872 Q98872 brachydanio

ALIGNMENTS

RESULT 1
ID NK25 RAT STANDARD; PRT; 318 AA.
AC Q35767;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE Homeobox protein Nkx-2.5 (Homeobox protein NK-2 homolog E) (rNKx-2.5).
GN Name=Nkx2-5; Synonyms=Cx, Nkx-2.5, Nkx2e;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RA Fu Y., Ruiz-Lozano P., Evans S.M.;
RT "A rat homeobox gene, rNKx-2.5, is a homologue of the tinman gene in Drosophila and is mainly expressed during heart development.";
RL Dev. Genes Evol. 207:352-358(1997).
CC -!- FUNCTION: Imlicated in commitment to and/or differentiation of the myocardial lineage (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Belongs to the NK-2 homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; AF006664; AAB62696.1; --
HSSP; P23441; 1FTT.
DR RGD; 620520; Csx.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeoboxdomain_like.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
KW Developmental protein; DNA-binding; Homeobox; Nuclear protein.
FT DOMAIN 41 107 Ala/Pro-rich.
FT DNA_BIND 137 196 Homeobox.
FT DOMAIN 206 280 Ala/Pro-rich.
SQ SEQUENCE 318 AA; 34336 MW; FCF845D4C41DA87C CRC64;

Query Match 66.7%; Score 68; DB 1; Length 318;
Best Local Similarity 63.6%; Pred No. 0.0073;
Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L20300; -; NOT_ANNOTATED_CDS.
 DR EMBL; AF091351; AAC97934.1; -;
 DR EMBL; X75415; CAA53169.1; -;
 DR HSSP; P23441; 1FTT.
 DR IntAct; P42582; -;
 DR TRANSFAC; T01675; -;
 DR MGD; MGI:97350; Nkx2-5.
 DR GO; GO:0005667; C:transcription factor complex; IDA.
 DR GO; GO:0003677; F:DNA binding; IDA.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0016564; F:transcriptional repressor activity; IDA.
 DR GO; GO:0000122; P:negative regulation of transcription from P...; IDA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR009057; Homeodomain_like.
 DR Pfam; PF00046; Homeobox; 1.
 DR PRINTS; PR00024; HOMEBOX.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX 1; 1.
 DR EMBL; X75415; CAA53169.1; -;
 KW Developmental protein; DNA-binding; Homeobox; Nuclear protein.
 FT DOMAIN 41 107 Ala/Pro-rich.
 FT DNA_BIND 137 196 Homeobox.
 FT DOMAIN 206 280 Ala/Pro-rich.
 FT CONFLICT 62 62 A -> R (in Ref. 4).
 FT CONFLICT 164 165 AP -> PA (in Ref. 1).
 FT CONFLICT 234 234 A -> T (in Ref. 3).
 FT CONFLICT 276 277 QP -> HA (in Ref. 1).
 SQ SEQUENCE 318 AA; 34162 MW; 6A3E7859F46D4FA6 CRC64;
 Query Match 63.7%; Score 65; DB 1; Length 318;
 Best Local Similarity 68.4%; Pred. No. 0.023;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 2 TKTPTSVKDIILSLPEQXRA 20
 DB 8 TPTPTSVKDIILNLEQQQRS 26
 RESULT 6
 Q925V3 PRELIMINARY; PRT; 318 AA.
 AC Q925V3;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Cardiac homeobox transcription factor.
 GN Name=Nkx2-5;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129;
 RA Tanaka M., Wechsler S.B., Lee I.W., Yamasaki N., Lawitts J., Izumo S.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 DR EMBL; AF083133; AAG38875.1; -;
 DR HSSP; P23441; 1FTT.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; P:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR Pfam; PF00046; Homeobox; 1.
 DR PRINTS; PR00024; HOMEBOX.
 DR ProDom; PD000010; Homeobox; 1.

DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX 1; 1.
 DR PROSITE; PS50071; HOMEBOX 2; 1.
 KW DNA-binding; Homeobox; Nuclear protein.
 SQ SEQUENCE 318 AA; 34145 MW; 6B8A22998C204EB6 CRC64;
 Query Match 63.7%; Score 65; DB 2; Length 318;
 Best Local Similarity 68.4%; Pred. No. 0.023;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 2 TKTPTSVKDIILSLPEQXRA 20
 DB 8 TPTPTSVKDIILNLEQQQRS 26
 RESULT 7
 Q98871 PRELIMINARY; PRT; 321 AA.
 ID Q98871
 AC Q98871;
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Homeodomain protein Nkx2.3.
 GN Name=nkx2.3;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AB; TISSUE=Whole embryo;
 RX MEDLINE=97111571; PubMed=8954740; DOI=10.1006/dbio.1996.0341;
 RA Lee K.H., Xu Q., Breitbart R.E.;
 RT "A new timman-related gene, nkx2.7, anticipates the expression of
 nkx2.5 and nkx2.3 in zebrafish heart and pharyngeal endoderm.";
 RL Dev. Biol. 180:722-731(1996).
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 DR EMBL; U66571; AAC05228.1; -;
 DR HSSP; P23441; 1FTT.
 DR ZFIN; ZDB-GENE-950415-178; nkx2.3.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; P:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR009057; Homeodomain_like.
 DR Pfam; PF00046; Homeobox; 1.
 DR PRINTS; PR00024; HOMEBOX.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX 1; 1.
 DR PROSITE; PS50071; HOMEBOX 2; 1.
 KW DNA-binding; Homeobox; Nuclear protein.
 SQ SEQUENCE 321 AA; 36092 MW; 3C45305A10029132 CRC64;
 Query Match 63.7%; Score 65; DB 2; Length 321;
 Best Local Similarity 72.2%; Pred. No. 0.024;
 Matches 13; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 2 TKTPTSVKDIILSLPEQXR 19
 DB 8 TSTPTSVKDIILKLEQQR 25
 RESULT 8
 Q91623 PRELIMINARY; PRT; 335 AA.
 ID Q91623
 AC Q91623;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Nkx-related protein.
 GN Name=NXKX-2.3;

OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 RN NCBI_TaxID=8355;
 RX SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RA MEDLINE=97108749; PubMed=8951070;
 RT Cleaver O.B., Patterson K.D., Krieg P.A.;
 RT "Overexpression of the tinman-related genes *XNkx-2.5* and *XNkx-2.3* in
 RT *Xenopus* embryos results in myocardial hyperplasia";
 RL Development 122:3549-3556(1996).
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 DR EMBL: U30318; AB49657.1; -.
 DR HSSP: P23441; 1FTT.
 DR GO: GO:0005634; C:nucleus; IEA.
 DR GO: GO:0003700; F:transcription factor activity; IEA.
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR Pfam: PF00046; Homeobox; 1.
 DR PRINTS: PR00024; HOMEBOX.
 DR ProDom: PD000010; Homeobox; 1.
 DR SMART: SM00389; HOX; 1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS50071; HOMEBOX_2; 1.
 DR DNA-binding; Homeobox; Nuclear protein.
 KW SEQUENCE 335 AA; 37402 MW; 14B8E5363C40ACF5 CRC64;
 SQ SEQUENCE 335 AA; 37402 MW; 14B8E5363C40ACF5 CRC64;
 Query Match 60.8%; Score 62; DB 2; Length 335;
 Best Local Similarity 66.7%; Pred. No. 0.079;
 Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 2 TKTFFSVKDIILSLPEQXR 19
 DB 8 TSTPFSVKDIILNLEQQAQ 25
 RESULT 9
 Q91861
 ID Q91861 PRELIMINARY; PRT; 335 AA.
 AC Q91861
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE Allele *XNkx-2.3a*, tinman homologue.
 GN Name=*XNkx-2.3*;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 RN NCBI_TaxID=8355;
 RX SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RA MEDLINE=96102830; PubMed=8582297;
 RA Evans S.M., Yan W., Murillo M.P., Ponce J., Papalopolu N.;
 RT "tinman, a Drosophila homeobox gene required for heart and visceral
 RT mesoderm specification, may be represented by a family of genes in
 RT vertebrates: *XNkx-2.3*, a second vertebrate homologue of tinman";
 RL Development 121:3889-3899(1995).
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 DR EMBL: L38675; AAA99180.1; -.
 DR HSSP: P23441; 1FTT.
 DR TRANSFAC: T04331; -.
 DR GO: GO:0005634; C:nucleus; IEA.
 DR GO: GO:0003700; F:transcription factor activity; IEA.
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR Pfam: PF00046; Homeobox; 1.
 DR PRINTS: PR00024; HOMEBOX.
 DR ProDom: PD000010; Homeobox; 1.
 DR SMART: SM00389; HOX; 1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS50071; HOMEBOX_2; 1.
 DR DNA-binding; Homeobox; Nuclear protein.
 KW SEQUENCE 329 AA; 36825 MW; DC1CE4E7275560EB CRC64;
 SQ SEQUENCE 329 AA; 36825 MW; DC1CE4E7275560EB CRC64;
 Query Match 59.8%; Score 61; DB 2; Length 329;
 Best Local Similarity 75.0%; Pred. No. 0.11;
 Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 TKTFFSVKDIILSLPEQ 17
 DB 8 TSTPFSVKDIILNLEQ 23
 RESULT 11
 Q91863
 ID Q91863 PRELIMINARY; PRT; 331 AA.
 AC Q91863
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE Allele *XNkx-2.3b*, tinman homologue.
 GN Name=*XNkx-2.3*;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.

KW DNA-binding; Homeobox; Nuclear protein.
 SQ SEQUENCE 335 AA; 37416 MW; 15338880FA87B090 CRC64;
 Query Match 60.8%; Score 62; DB 2; Length 335;
 Best Local Similarity 66.7%; Pred. No. 0.079;
 Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 2 TKTFFSVKDIILSLPEQXR 19
 DB 8 TSTPFSVKDIILNLEQQAQ 25
 RESULT 10
 Q91862
 ID Q91862 PRELIMINARY; PRT; 329 AA.
 AC Q91862
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE Allele *XNkx-2.3bi*, tinman homologue.
 GN Name=*XNkx-2.3*;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 RN NCBI_TaxID=8355;
 RX SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RA MEDLINE=96102830; PubMed=8582297;
 RA Evans S.M., Yan W., Murillo M.P., Ponce J., Papalopolu N.;
 RT "tinman, a Drosophila homeobox gene required for heart and visceral
 RT mesoderm specification, may be represented by a family of genes in
 RT vertebrates: *XNkx-2.3*, a second vertebrate homologue of tinman";
 RL Development 121:3889-3899(1995).
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 DR EMBL: L38675; AAA99180.1; -.
 DR HSSP: P23441; 1FTT.
 DR TRANSFAC: T04331; -.
 DR GO: GO:0005634; C:nucleus; IEA.
 DR GO: GO:0003700; F:transcription factor activity; IEA.
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR Pfam: PF00046; Homeobox; 1.
 DR PRINTS: PR00024; HOMEBOX.
 DR ProDom: PD000010; Homeobox; 1.
 DR SMART: SM00389; HOX; 1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS50071; HOMEBOX_2; 1.
 DR DNA-binding; Homeobox; Nuclear protein.
 KW SEQUENCE 329 AA; 36825 MW; DC1CE4E7275560EB CRC64;
 SQ SEQUENCE 329 AA; 36825 MW; DC1CE4E7275560EB CRC64;
 Query Match 59.8%; Score 61; DB 2; Length 329;
 Best Local Similarity 75.0%; Pred. No. 0.11;
 Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 TKTFFSVKDIILSLPEQ 17
 DB 8 TSTPFSVKDIILNLEQ 23
 RESULT 11
 Q91863
 ID Q91863 PRELIMINARY; PRT; 331 AA.
 AC Q91863
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE Allele *XNkx-2.3b*, tinman homologue.
 GN Name=*XNkx-2.3*;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.


```

OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=96102830; PubMed=8582297;
RA Evans S.M., Yan W., Muxillo M.P., Ponce J., Papalopulu N.;
RT "timan, a Drosophila homeobox gene required for heart and visceral
RT mesoderm specification, may be represented by a family of genes in
RT vertebrates: NKX-2.3, a second vertebrate homologue of tinman.";
RL Development 121:3889-3899(1995).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR EMBL; L38676; AA99181.1; -
DR HSP; P23441; 1FTT.
DR TRANSFAC; T04332; -
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR Pfam; PF00046; Homeobox.1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox.1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 331 AA; 37130 MW; 1AP54E3C89703DCD CRC64;

Query Match 59.8%; Score 61; DB 2; Length 331;
Best Local Similarity 75.0%; Pred. No. 0.12;
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 TKTTPSVKDIILSLPEQ 17
DB 8 TSTPFSVKDIILNLEQ 23

RESULT 12
NK23_HUMAN
ID NK23_HUMAN STANDARD; PRT; 364 AA.
AC Q8TAU0; Q9NYS6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Homeobox protein NKX-2.3 (Homeobox protein NKX-2.3) (Homeobox protein
DE NK-2 homolog C).
GN Name=NKX2-3; Synonyms=NKX23, NKX2C;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ramsay H.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas, and Spleen;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner B.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.K., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny N.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Heltan E., Kettunen M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,

```

```

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 1-211 FROM N.A.
RA Wang C.-C., Brodnicki T., Harvey R.P.;
RT "Homeobox gene NKX2-3 is critical for normal development of the gut,
RT spleen and gut-associated lymphoid tissue.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Transcription factor (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: Belongs to the NK-2 homeobox family.
CC -1- SIMILARITY: Contains 1 homeobox domain.
CC -1- CAUTION: Ref.2 sequence differs from that shown due to a potential
CC the sequence of residues 274-302 are lacking from the translated
CC cDNA while this region does not seem to correspond to an exon.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AL353719; -; NOT ANNOTATED CDS.
DR EMBL; BC025788; AAH25788.1; ALT_FRAME.
DR EMBL; AF229631; AAF44651.1; -
DR EMBL; AF229630; AAF44651.1; JOINED.
DR HSP; P23441; 1FTT.
DR TRANSFAC; T04341; -
DR Genew; HGNC:7836; NKX2-3.
DR MIM; 606727; -
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeobox_main_like.
DR Pfam; PF00046; Homeobox.1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox.1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein; Transcription regulation.
FT DNA BIND 148 207 Homeobox.
FT DOMAIN 219 223 Poly-Pro.
FT DOMAIN 271 286 Poly-Ala.
FT DOMAIN 298 305 Poly-Gly.
FT SEQUENCE 364 AA; 38405 MW; 10A279CFA3C9A6DB CRC64;

Query Match 59.8%; Score 61; DB 1; Length 364;
Best Local Similarity 75.0%; Pred. No. 0.13;
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 TKTTPSVKDIILSLPEQ 17
DB 8 TSTPFSVKDIILNLEQ 23

RESULT 13
Q9HBZ5
ID Q9HBZ5 PRELIMINARY; PRT; 86 AA.
AC Q9HBZ5;
DT 01-MAR-2001 (TRENDELrel. 16, Created)
DT 01-MAR-2001 (TRENDELrel. 16, Last sequence update)
DT 01-MAR-2001 (TRENDELrel. 16, Last annotation update)
DE Transcription factor (Fragment).
GN Name=NKX2B;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```



```

RP SEQUENCE FROM N.A.
RX MEDLINE=20279852; PubMed=10818213;
RA Wang C.-C., Brodnicki T., Copeland N.G., Jenkins N.A., Harvey R.P.;
RT "Conserved linkage of NK-2 homeobox gene pairs Nkx2-2/2-4 and Nkx2-
RL 1/2-9 in mammals.";
DR Mamm. Genome 11:466-468(2000).
FT EMBL; AF184045; AAG16976.1; -.
SQ SEQUENCE 86 AA; 9246 MW; F4CCAD21324BE9D1 CRC64;

Query Match 58.8%; Score 60; DB 2; Length 86;
Best Local Similarity 59.1%; Pred. No. 0.041;
Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 TKTGPSVKDILSLPQXRAXGA 23
DB 6 TKTGPSVKDILSLPQXRAXGA 27

RESULT 14
Q9JKR0 PRELIMINARY; PRT; 86 AA.
ID Q9JKR0;
AC Q9JKR0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Homeobox protein Nkx2-2 (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SV;
RA Wang C.-C., Brodnicki T., Copeland N.G., Jenkins N.A., Harvey R.P.;
RT "Conserved Linkage of NK-2 Homeobox Gene Pairs Nkx2-2/2-4 and Nkx2-
RL 1/2-9 in mammals.";
RL Mamm. Genome 0:0-0(2000).
DR EMBL; AF229632; AAF44652.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
KW Nuclear protein.
FT NON_TER 86
SQ SEQUENCE 86 AA; 9191 MW; CED0E9A7E933D8D8 CRC64;

Query Match 58.8%; Score 60; DB 2; Length 86;
Best Local Similarity 59.1%; Pred. No. 0.041;
Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 TKTGPSVKDILSLPQXRAXGA 23
DB 6 TKTGPSVKDILSLPQXRAXGA 27

RESULT 15
Q6DGE7 PRELIMINARY; PRT; 143 AA.
ID Q6DGE7;
AC Q6DGE7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Singapore local strain; TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.N., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

```

```

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Sklisswood J., Schmutz J., Myers R.W., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalut D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Singapore local strain; TISSUE=Embryo;
RA Strausberg R.;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC076402; AAH76402.1; -.
KW Hypothetical protein.
SQ SEQUENCE 143 AA; 15383 MW; 54EB7F50FD8C573 CRC64;

Query Match 58.8%; Score 60; DB 2; Length 143;
Best Local Similarity 59.1%; Pred. No. 0.071;
Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 TKTGPSVKDILSLPQXRAXGA 23
DB 6 TKTGPSVKDILSLPQXRAXGA 27

RESULT 16
Q8BR9 PRELIMINARY; PRT; 248 AA.
ID Q8BR9;
AC Q8BR9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 10 days neonate cortex cDNA, RIKEN full-length enriched
DE library, clone:A83008A18 product:MK2 transcription factor related,
DE locus 2 (Drosophila), full insert sequence.
CN Name=Nkx2-2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=9927253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]

```


RP
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes."
RL Genome Res. 10:1617-1630(2000).
RN [15]

RP
RC SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujikawa S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsunaga S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RA "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer."
RL Genome Res. 10:1757-1771(2000).
RN [16]

RP
RC SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Cortex;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Nunazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh K., Sakai K., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Taya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR EMBL; AK043561; BAC1582.1; -.
DR HSSP; P22808; INK3.
DR MGD; MGI:97347; Nkx2-2.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0003700; F:transcription factor activity; IDA.
DR GO; GO:0007399; P:neurogenesis; IDA.
DR GO; GO:0045597; P:positive regulation of cell differentiation; IMP.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeodomain_like.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 248 AA; 27798 MW; 00F062658AD0702D CRC64;

Query Match 58.8%; Score 60; DB 2; Length 248;
Best Local Similarity 59.1%; Pred. No. 0.13;
Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 TKTFFSVKDIILSLPEQXRAXGA 23
||| ||||| |||
Db 6 TKTGFSVKDILDLPTNDEGS 27

RESULT 17
NK22 BRARE STANDARD; PRT; 269 AA.
AC Q90481;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Homeobox protein Nkx-2.2 (Homeobox protein Nkx-2 homolog B).
GN Name=nkx2.2; Synonyms=nkx2.2, nkx2-2;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95324401; PubMed=7600991;
RA Barth K.A., Wilson S.W.;
RT "Expression of zebrafish nkx2.2 is influenced by sonic
RT hedgehog/vertebrate hedgehog-1 and demarcates a zone of neuronal
RT differentiation in the embryonic forebrain."
RL Development 121:1755-1768(1995).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- TISSUE SPECIFICITY: Expressed in a continuous narrow band of cells
CC along a boundary zone demarcating the location at which two of the
CC earliest nuclei in the brain differentiate.
CC -!- SIMILARITY: Belongs to the NK-2 homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; X85977; CAA59967.1; -.
DR PIR; I50504; I50504.
DR HSSP; P22808; INK3.
DR ZFIN; ZDB-GENE-980526-403; nkx2.2.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeodomain_like.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW Developmental protein; DNA-binding; Homeobox; Nuclear protein.
FT DNA BIND 125 184
SO SEQUENCE 269 AA; 30306 MW; CF006285CDD33D63 CRC64;

Query Match 58.8%; Score 60; DB 1; Length 269;
Best Local Similarity 59.1%; Pred. No. 0.14;
Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 TKTFFSVKDIILSLPEQXRAXGA 23
||| ||||| |||
Db 6 TKTGFSVKDILDLPTNDEGS 27

RESULT 18
Q98873 PRELIMINARY; PRT; 269 AA.
AC Q98873;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Homeodomain protein Nkx2.7.
GN Name=nkx2.7;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Whole embryo;


```

RX MEDLINE=9711517; PubMed=8954740; DOI=10.1006/dbio.1996.0341;
RA Lee K.H., Xu Q., Breitbart R.E.;
RT "A new timan-related gene, nkx2.7, anticipates the expression of
RT nkx2.5 and nkx2.3 in zebrafish heart and pharyngeal endoderm.";
RL Dev. Biol. 180:722-731(1996).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR EMBL; U66573; AAC05230.1; -
DR HSP; P23441; 1FTT
DR ZFIN; ZDB-GENE-990415-179; nkx2.7.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeodomain_like.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 269 AA; 30788 MW; 83565C5A4667DEA CRC64;

Query Match 58.8%; Score 60; DB 2; Length 269;
Best Local Similarity 75.0%; Pred. No. 0.14;
Matches 12; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 TKTSPSVKDIILSLPEQ 17
DB 7 TSTPFSVKDILKLEQQ 22

RESULT 19
ID Q6AXL1 PRELIMINARY; PRT; 269 AA.
AC Q6AXL1;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE NK2 transcription factor related 7.
GN Name=nkx2.7;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Singapore local strain; TISSUE=Embryo;
RX PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo A.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywicki M.I., Skalska U., Smullus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Singapore local strain; TISSUE=Embryo;

```

```

RA Director MGC Project;
RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR EMBL; BC079493; AAH79493.1; -
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeodomain_like.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 269 AA; 30801 MW; 24E00798CC7506CO CRC64;

Query Match 58.8%; Score 60; DB 2; Length 269;
Best Local Similarity 75.0%; Pred. No. 0.14;
Matches 12; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 TKTSPSVKDIILSLPEQ 17
DB 7 TSTPFSVKDILKLEQQ 22

RESULT 20
ID NK22 HUMAN
AC O95096;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Homeobox protein NKx-2.2 (Homeobox protein NK-2 homolog B).
GN Name=NKX2-2; Synonyms=NKX2.2, NKX2B;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98366900; PubMed=9703340;
RX Futaba H., Horikawa Y., Iwasaki N., Hara M., Sussel L., le Beau M.M.,
RX Davis E.M., Ogata M., Iwamoto Y., German M.S., Bell G.I.;
RA "Beta-cell transcription factors and diabetes: mutations in the coding
RT region of the BETA2/NeuroD1 (NEUROD1) and Nkx2.2 (NKX2B) genes are not
RT associated with maturity-onset diabetes of the young in Japanese.";
RL Diabetes 47:1356-1358(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052; DOI=10.1038/414865a;
RA Deloukas P., Matthews L.H., Ashurst J.L., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beare L.M., Beare D.W.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leivasalio M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLeay K., Nickerson A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Peck A.I.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Pathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Suleston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,

```


RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
CC -!- FUNCTION: May be involved in specifying diencephalic neuromeric
CC boundaries, and in controlling the expression of genes that play a
CC role in axonal guidance (By similarity).
CC -!- SUBUNIT: Interacts with OLIG2 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- DOMAIN: The homeobox domain is essential for interaction with OLIG2
CC (By similarity).
CC -!- SIMILARITY: Belongs to the NK-2 homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF019415; AAC83132.1; -.
CC EMBL; AF019414; AAC83132.1; JOINED.
CC EMBL; ALI33325; CAC00770.1; -.
CC HSSP; P22808; INK3.
CC TRANSFAC; T04337; -.
CC GENEX; HGNC:7835; NKX2-2.
CC MIM; 604612; -.
CC GO; GO:0007420; P:brain development; TAS.
CC InterPro; IPR001356; Homeobox.
CC InterPro; IPR009057; Homeobox-like.
CC Pfam; PF00046; Homeobox; 1.
CC PRINTS; PRO0024; HOMEBOX.
CC ProDom; PD000010; Homeobox; 1.
CC SMART; SM00389; HOX; 1.
CC MIM; 604612; -.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS00071; HOMEBOX_2; 1.
CC Developmental protein; DNA-binding; Homeobox; Nuclear protein.
KW DNA BIND 128 187 Homeobox.
FT DNA BIND 128 187 Homeobox.
SQ SEQUENCE 273 AA; 30132 MW; 91BDA220BAFA63D1 CRC64;

Query Match 58.8%; Score 60; DB 1; Length 273;
Best Local Similarity 59.1%; Pred. No. 0.14;
Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 TKTGPSVKDILSLPEQXRAXGA 23
DB 6 TKTGPSVKDILSLPEQXRAXGA 27

RESULT 21
NK22_MESAU
ID NK22_MESAU STANDARD; PRT; 273 AA.
AC P43697;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Homeobox protein Nkx-2.2 (Homeobox protein NK-2 homolog B).
GN Name=NKX2-2; Synonyms=NKX-2.2, NKX2B;
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Pancratic islets;
RX MEDLINE=95083670; PubMed=7991607;
RA Rudnick A., Ling T.Y., Odagiri H., Rutter W.J., German M.S.;
RT "Pancratic beta cells express a diverse set of homeobox genes.";

Proc. Natl. Acad. Sci. U.S.A. 91:12203-12207(1994).
-!- FUNCTION: May be involved in specifying diencephalic neuromeric
CC boundaries, and in controlling the expression of genes that play a
CC role in axonal guidance (By similarity).
CC -!- SUBUNIT: Interacts with OLIG2 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- DOMAIN: The homeobox domain is essential for interaction with OLIG2
CC (By similarity).
CC -!- SIMILARITY: Belongs to the NK-2 homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X81408; CAA57165.1; -.
CC PIR; I48187; I48187.
CC HSSP; P22808; INK3.
CC TRANSFAC; T04265; -.
CC InterPro; IPR001356; Homeobox.
CC InterPro; IPR009057; Homeobox-like.
CC Pfam; PF00046; Homeobox; 1.
CC PRINTS; PRO0024; HOMEBOX.
CC ProDom; PD000010; Homeobox; 1.
CC SMART; SM00389; HOX; 1.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS00071; HOMEBOX_2; 1.
CC Developmental protein; DNA-binding; Homeobox; Nuclear protein.
KW DNA BIND 128 187 Homeobox.
FT DNA BIND 128 187 Homeobox.
SQ SEQUENCE 273 AA; 30064 MW; 3269F6F699E922A2 CRC64;

Query Match 58.8%; Score 60; DB 1; Length 273;
Best Local Similarity 59.1%; Pred. No. 0.14;
Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 TKTGPSVKDILSLPEQXRAXGA 23
DB 6 TKTGPSVKDILSLPEQXRAXGA 27

RESULT 22
NK22_MOUSE
ID NK22_MOUSE STANDARD; PRT; 273 AA.
AC P42586;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Homeobox protein Nkx-2.2 (Homeobox protein NK-2 homolog B).
GN Name=NKx2-2; Synonyms=NKx-2.2, Nkx2b;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=95025968; PubMed=7939711;
RA Rubenstein J.L., Martinez S., Shimamura K., Puelles L.;
RT "The embryonic vertebrate forebrain: the prosomeric model.";
RL Science 266:578-580(1994).
RN [2]
RN SEQUENCE FROM N.A.
RX PubMed=8654958; DOI=0378111995007458;
RA Hartigan D.J., Rubenstein J.L.;
RT "The cDNA sequence of murine Nkx-2.2.";
RL Gene 168:271-272(1996).
RN [3]
RN SEQUENCE OF 112-196 FROM N.A.
RX MEDLINE=92153416; PubMed=1346742;
RA Price M., Lazzaro D., Pohl T., Mattei M.-G., Ruether U., Olivo J.-C.,


```

RA Duboule D., di Lauro R.;
RT "Regional expression of the homeobox gene Nkx-2.2 in the developing
RT mammalian forebrain";
RL Neuron 8:241-255(1992).
RN [4]
RP INTERACTION WITH OLIG2, AND SUBCELLULAR LOCATION.
RX PubMed=14575334; DOI=23/29/9547;
RA Sun T., Dong H., Wu L., Kane M., Rowitch D.H., Stiles C.D.;
RT "Cross-repressive interaction of the Olig2 and Nkx2.2 transcription
RT factors in developing neural tube associated with formation of a
RT specific physical complex.";
RL J. Neurosci. 23:9547-9556(2003).
CC -!- FUNCTION: May be involved in specifying diencephalic neuromeric
CC boundaries, and in controlling the expression of genes that play a
CC role in axonal guidance.
CC -!- SUBUNIT: Interacts with OLIG2.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Expressed in restricted areas of the
CC developing CNS: the hindbrain and forebrain, and pancreas.
CC -!- DEVELOPMENTAL STAGE: Expressed during embryogenesis.
CC -!- DOMAIN: The homeodomain is essential for interaction with OLIG2.
CC -!- SIMILARITY: Belongs to the NK-2 homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U31566; AAA79303.1; -.
CC EMBL; S83259; -. NOT_ANNOTATED_CDS.
CC F1R; J4634; J4634.
CC HSSP; P22808; INK3.
CC TRANSFAC; T02384; -.
CC MGD; MGI:97347; Nkx2-2.
CC DR GO; GO:0005634; C:nucleus; IDA.
CC DR GO; GO:0005515; F:protein binding; IPI.
CC DR GO; GO:0003700; F:transcription factor activity; IDA.
CC DR GO; GO:0007399; P:neurogenesis; IDA.
CC DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
CC DR InterPro; IPR001356; Homeobox.
CC DR InterPro; IPR009057; Homeobox.
CC DR Pfam; PF00046; Homeobox; 1.
CC DR PRINTS; PR00024; HOMEBOX.
CC DR ProDom; PD000010; Homeobox; 1.
CC DR SMART; SM00389; HOX; 1.
CC DR PROSITE; PS00027; HOMEBOX_1; 1.
CC DR PROSITE; PS50071; HOMEBOX_2; 1.
CC DR Developmental protein; DNA-binding; Homeobox; Nuclear protein.
CC FT DNA_BIND 128 187 Homeobox.
CC FT SEQUENCE 273 AA; 30125 MW; 1F3DBB8127667F19 CRC64;
CC
CC Query Match 58.8%; Score 60; DB 1; Length 273;
CC Best Local Similarity 59.1%; Pred. No. 0.14;
CC Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
CC
CC QY 2 TKTSPFVKDILDLPEQXRAXGA 23
CC ||| ||||| |||
CC Db 6 TKTGFSVKDILDLPTNDRDGS 27
CC
CC RESULT 23
CC NK25_CHICK
CC ID NK25_CHICK STANDARD; PRT; 294 AA.
CC AC Q90788;
CC DT 15-DEC-1998 (Rel. 37, Created)
CC DT 15-DEC-1998 (Rel. 37, Last sequence update)
CC DT 05-JUL-2004 (Rel. 44, Last annotation update)
CC DE Homeobox protein Nkx-2.5 (Homeobox protein NK-2 homolog E) (CNKX-2.5).
CC GN Name=NKX-2.5; Synonyms=NKX2E;

```

```

OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96125161; PubMed=8575320;
RA Schultheiss T.M., Xydas S., Lassar A.B.;
RT "Induction of avian cardiac myogenesis by anterior endoderm.";
RL Development 121:4203-4214(1995).
CC -!- FUNCTION: Implicated in commitment to and/or differentiation of
CC the myocardial lineage.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Belongs to the NK-2 homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X91838; CAA62955.1; -.
CC HSSP; P23441; 1FTT.
CC TRANSFAC; T04334; -.
CC InterPro; IPR001356; Homeobox.
CC DR InterPro; IPR009057; Homeobox.
CC DR Pfam; PF00046; Homeobox; 1.
CC DR PRINTS; PR00024; HOMEBOX.
CC DR ProDom; PD000010; Homeobox; 1.
CC DR SMART; SM00389; HOX; 1.
CC DR PROSITE; PS00027; HOMEBOX_1; 1.
CC DR PROSITE; PS50071; HOMEBOX_2; 1.
CC DR Developmental protein; DNA-binding; Homeobox; Nuclear protein.
CC FT DOMAIN 21 24 Poly-Gln.
CC FT DNA_BIND 119 178 Homeobox.
CC FT SEQUENCE 294 AA; 33072 MW; E963373B2C074A32 CRC64;
CC
CC Query Match 58.8%; Score 60; DB 1; Length 294;
CC Best Local Similarity 75.0%; Pred. No. 0.15;
CC Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
CC
CC QY 2 TKTSPFVKDILSLPEQ 17
CC ||| ||||| |||
CC Db 7 TTTTSPVKDILNLSEQQ 22
CC
CC RESULT 24
CC OL3005
CC ID OL3005 PRELIMINARY; PRT; 193 AA.
CC AC OL3005;
CC DT 01-JUL-1997 (TRENBLrel. 04, Created)
CC DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
CC DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
CC DE Nkx2-8 protein.
CC GN Name=Nkx2-8;
CC OS Gallus gallus (Chicken).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
CC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=White Leghorn;
RA Brand T., Andre B., Schneider A., Arnold H.H.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ database.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC EMBL; Y11949; CAA72693.1; -.
CC DR HSSP; P23441; 1FTT.
CC TRANSFAC; T04340; -.

```


DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR009057; Homeobox.
 DR Pfam; PF00046; Homeobox; 1.
 DR PRINTS; PR00024; HOMEBOX.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS00071; HOMEBOX_2; 1.
 DR DNA-binding; Homeobox; Nuclear protein.
 KW SEQUENCE 193 AA; 21717 MW; 43A46C374C8E2A9C CRC64;
 SQ

Query Match 58.3%; Score 59.5; DB 2; Length 193;
 Best Local Similarity 75.0%; Pred. No. 0.12;
 Matches 15; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 4 TPFSVKDILSLPEQXRAXGA 23
 DB 4 TPFSVEDILSL-EQSSAPGA 22

RESULT 25
 OI3053 PRELIMINARY; PRT; 193 AA.
 AC OI3053;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE NKX2.8 protein.
 GN Name=NKX2.8;
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Boettger T., Stein S., Kessel M.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 DR EMBL; Y10655; CAA71665.1; -.
 DR HSP; P23441; IFTT.
 DR TRANSFAC; T04340; -.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR009057; Homeobox.
 DR Pfam; PF00046; Homeobox; 1.
 DR ProDom; PD000010; Homeobox; 1.
 DR PRINTS; PR00024; HOMEBOX.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS00071; HOMEBOX_2; 1.
 DR DNA-binding; Homeobox; Nuclear protein.
 KW SEQUENCE 193 AA; 21917 MW; 8B8BBE751C196A4 CRC64;
 SQ

Query Match 58.3%; Score 59.5; DB 2; Length 193;
 Best Local Similarity 75.0%; Pred. No. 0.12;
 Matches 15; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 4 TPFSVKDILSLPEQXRAXGA 23
 DB 4 TPFSVEDILSL-EQSSAPGA 22

RESULT 26
 Q6DN74 PRELIMINARY; PRT; 54 AA.
 ID Q6DN74;
 AC Q6DN74;
 DT 25-OCT-2004 (TREMBLrel. 28, Created)

DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)
 DE Homeodomain protein Nkx2-10 (Fragment).
 GN Name=Nkx2-10;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=89355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Allen B.G., Weeks D.L.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY654625; AAT73698.1; -.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR Nuclear protein.
 FT NON TER 54 54
 SQ SEQUENCE 54 AA; 5716 MW; 5A80F5F8C831572 CRC64;
 Query Match 56.9%; Score 58; DB 2; Length 54;
 Best Local Similarity 57.1%; Pred. No. 0.055;
 Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 TKTPFSVKDILSLPEQXRAXG 22
 DB 7 TSTPFSVKDILRLLEQLQSDG 27

RESULT 27
 Q9W6H4 PRELIMINARY; PRT; 263 AA.
 ID Q9W6H4;
 AC Q9W6H4;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Homeodomain protein XNkx2-10.
 GN Name=XNkx2-10;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=89355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=98284450; PubMed=9621430;
 RX DOI=10.1002/(SICI)1520-6408(1998)22:3<230::AID-DVG5>3.3.CO;2-9;
 RA Newman C.S., Krieg P.A.;
 RT "tinman-related genes expressed during heart development in Xenopus.";
 RL Dev. Genet. 22:230-238(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20171065; PubMed=10704867; DOI=10.1016/S0925-4773(99)00291-9;
 RA Newman C.S., Reecy J., Grow M.W., Ni K., Boettger T., Kessel M.,
 RA Schwartz R.J., Krieg P.A.;
 RT "Transient cardiac expression of the tinman-family homeobox gene,
 RT XNkx2-10.";
 RL Mech. Dev. 91:369-373(2000).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 DR EMBL; AF127224; AAD30270.1; -.
 DR HSP; P23441; IFTT.
 DR TRANSFAC; T04335; -.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR009057; Homeobox.
 DR Pfam; PF00046; Homeobox; 1.
 DR PRINTS; PR00024; HOMEBOX.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS00071; HOMEBOX_2; 1.
 KW DNA-binding; Homeobox; Nuclear protein.

SQ SEQUENCE 263 AA; 29273 MW; F9258527A0D7207A CRC64;
 Query Match 56.9%; Score 58; DB 2; Length 263;
 Best Local Similarity 57.1%; Pred. No. 0.29;
 Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 QY 2 TKTSPSVKDIILSLPEQXRXG 22
 Db 7 TSTPSPVKDILRLQEQSDG 27
 RESULT 28
 O96377
 ID O96377 PRELIMINARY; PRT; 312 AA.
 AC O96377;
 DT 01-MAY-1999 (TReMBLrel. 10, Created)
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
 DT 01-MAY-2004 (TReMBLrel. 26, Last annotation update)
 DE Homeodomain protein.
 GN Name=NK2-2;
 OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 OX NCBI_TaxID=7739;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98237856; PubMed=9569351;
 RA Holland L.Z., Venkatesh T.V., Gorlin A., Bodmer R., Holland N.D.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 DR EMBL; AF032999; AAD01958.1; -.
 DR HSP; P22808; INK3.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; P:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR009057; Homeodomain_like.
 DR Pfam; PF00046; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR ProDom; PD00010; Homeobox; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS00071; HOMEBOX_2; 1.
 DR DNA-binding; Homeobox; Nuclear protein.
 KW DNA-binding; Homeobox; Nuclear protein.
 FT NON TER 394
 SQ SEQUENCE 312 AA; 34491 MW; 4A338D2ADC82AF63 CRC64;
 Query Match 55.9%; Score 57; DB 2; Length 312;
 Best Local Similarity 61.1%; Pred. No. 0.51;
 Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 2 TKTSPSVKDIILSLPEQXRX 19
 Db 9 SKTSFTVKDILDMPEAKR 26
 RESULT 29
 O6QB01
 ID O6QB01 PRELIMINARY; PRT; 394 AA.
 AC O6QB01;
 DT 05-JUL-2004 (TReMBLrel. 27, Created)
 DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
 DE NK2.2 (Fragment).
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;

OC Strongylocentrotus.
 OX NCBI_TaxID=7668;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=15081364; DOI=10.1016/j.ydbio.2004.01.023;
 RA Takacs C.M., Amore G., Oliveri P., Pouska A.J., Wang D., Burke R.D.,
 RA Peterson K.J.;
 RL "Expression of an NK2 homeodomain gene in the apical ectoderm defines
 a new territory in the early sea urchin embryo.";
 RL Dev. Biol. 269:152-164(2004).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 DR EMBL; AY549449; AAS58444.1; -.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; P:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR009057; Homeodomain_like.
 DR Pfam; PF00046; Homeobox; 1.
 DR PRINTS; PR00024; HOMEBOX.
 DR ProDom; PD00010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS00071; HOMEBOX_2; 1.
 DR DNA-binding; Homeobox; Nuclear protein.
 KW DNA-binding; Homeobox; Nuclear protein.
 FT NON TER 394
 SQ SEQUENCE 394 AA; 41820 MW; AAB60A92AE6956E8 CRC64;
 Query Match 55.9%; Score 57; DB 2; Length 394;
 Best Local Similarity 61.9%; Pred. No. 0.66;
 Matches 13; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 QY 3 KTFPSVKDILSLPEQXRXGA 23
 Db 73 KTFPSVKDILDLPHAAVAAA 93
 RESULT 30
 Q9DDU7
 ID Q9DDU7 PRELIMINARY; PRT; 77 AA.
 AC Q9DDU7;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
 DE Nkx2-5 (Fragment).
 GN Name=Nkx2-5;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20530465; PubMed=11076677; DOI=10.1006/dbio.2000.9891;
 RA Sparrow D.B., Cai C., Kotecha S., Latinkic B., Cooper B., Towers N.,
 RA Evans S.M., Mohun T.J.;
 RL "Regulation of the tinman homologues in Xenopus embryos.";
 RL Dev. Biol. 227:65-79(2000).
 DR EMBL; AF283102; AAG42363.1; -.
 DR NON TER 77
 SQ SEQUENCE 77 AA; 8401 MW; F1605900F348FFFD CRC64;
 Query Match 54.9%; Score 56; DB 2; Length 77;
 Best Local Similarity 73.3%; Pred. No. 0.17;
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 TKTSPSVKDIILSLPE 16
 Db 7 TSTPSPVKDILNLEQ 21
 RESULT 31
 NK25_XENLA
 ID NK25_XENLA STANDARD; PRT; 299 AA.

Q6DUE1
ID Q6DUE1 PRELIMINARY; PRT; 300 AA.
AC Q6DUE1
DT 25-OCT-2004 (TRENBLrel. 28, Created)
DT 25-OCT-2004 (TRENBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TRENBLrel. 28, Last annotation update)
DE NK-2 class homeodomain protein.
GN Name=Nxx2.5;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Brown D.D., Martz S.N., Binder O., Price B.M.J., Smith J.C.,
RA Conlon F.L.;
RA SUBMITTED (JUN-2004) to the EMBL/GenBank/DBJ databases.
RC 1- SUBCELLULAR LOCATION: Nuclear (By similarity).
RL EMBL; AY644403; AT66927.1; -
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeodomain_like.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 2; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 300 AA; 34117 MW; 6359A5FF7C7316PF CRC64;
Query Match 54.9%; Score 56; DB 2; Length 300;
Best Local Similarity 73.3%; Pred. No. 0.73;
Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 2 TKTPTFSVKDILSLPE 16
DB 7 TSTPTFSVKDILNLEQ 21
RESULT 34
Q6UKJ7
ID Q6UKJ7 PRELIMINARY; PRT; 252 AA.
AC Q6UKJ7
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Timman (Fragment).
GN Name=tim;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=F-775F, S-438S, S-968F, S-565F, S-483F, US-255F, S-26F, S-510S,
RC S-521S, S-549S, and P-96S;
RX PubMed=15126403; DOI=10.1534/genetics.166.4.1845;
RA Balakirev E.S., Ayala F.J.;
RT "Nucleotide Variation in the tinman and bagpipe Homeobox Genes of
RT Drosophila melanogaster."
RL Genetics 166:1845-1856(2004).
DR EMBL; AY368098; AAQ75408.1; -
DR EMBL; AY368099; AAQ75409.1; -
DR EMBL; AY368100; AAQ75410.1; -
DR EMBL; AY368101; AAQ75411.1; -
DR EMBL; AY368102; AAQ75412.1; -
DR EMBL; AY368103; AAQ75413.1; -
DR EMBL; AY368104; AAQ75414.1; -
Q6DUE1
ID Q6DUE1 PRELIMINARY; PRT; 300 AA.
AC Q6DUE1
DT 25-OCT-2004 (TRENBLrel. 28, Created)
DT 25-OCT-2004 (TRENBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TRENBLrel. 28, Last annotation update)
DE NK-2 class homeodomain protein.
GN Name=Nxx2.5;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Brown D.D., Martz S.N., Binder O., Price B.M.J., Smith J.C.,
RA Conlon F.L.;
RA SUBMITTED (JUN-2004) to the EMBL/GenBank/DBJ databases.
RC 1- SUBCELLULAR LOCATION: Nuclear (By similarity).
RL EMBL; AY644403; AT66927.1; -
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeodomain_like.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 2; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 300 AA; 34117 MW; 6359A5FF7C7316PF CRC64;
Query Match 54.9%; Score 56; DB 2; Length 300;
Best Local Similarity 73.3%; Pred. No. 0.73;
Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 2 TKTPTFSVKDILSLPE 16
DB 7 TSTPTFSVKDILNLEQ 21
RESULT 34
Q6UKJ7
ID Q6UKJ7 PRELIMINARY; PRT; 252 AA.
AC Q6UKJ7
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Timman (Fragment).
GN Name=tim;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=F-775F, S-438S, S-968F, S-565F, S-483F, US-255F, S-26F, S-510S,
RC S-521S, S-549S, and P-96S;
RX PubMed=15126403; DOI=10.1534/genetics.166.4.1845;
RA Balakirev E.S., Ayala F.J.;
RT "Nucleotide Variation in the tinman and bagpipe Homeobox Genes of
RT Drosophila melanogaster."
RL Genetics 166:1845-1856(2004).
DR EMBL; AY368098; AAQ75408.1; -
DR EMBL; AY368099; AAQ75409.1; -
DR EMBL; AY368100; AAQ75410.1; -
DR EMBL; AY368101; AAQ75411.1; -
DR EMBL; AY368102; AAQ75412.1; -
DR EMBL; AY368103; AAQ75413.1; -
DR EMBL; AY368104; AAQ75414.1; -
Q6DUE1
ID Q6DUE1 PRELIMINARY; PRT; 300 AA.
AC Q6DUE1
DT 25-OCT-2004 (TRENBLrel. 28, Created)
DT 25-OCT-2004 (TRENBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TRENBLrel. 28, Last annotation update)
DE NK-2 class homeodomain protein.
GN Name=Nxx2.5;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Brown D.D., Martz S.N., Binder O., Price B.M.J., Smith J.C.,
RA Conlon F.L.;
RA SUBMITTED (JUN-2004) to the EMBL/GenBank/DBJ databases.
RC 1- SUBCELLULAR LOCATION: Nuclear (By similarity).
RL EMBL; AY644403; AT66927.1; -
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR009057; Homeodomain_like.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 2; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 300 AA; 34117 MW; 6359A5FF7C7316PF CRC64;
Query Match 54.9%; Score 56; DB 2; Length 300;
Best Local Similarity 73.3%; Pred. No. 0.73;
Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 2 TKTPTFSVKDILSLPE 16
DB 7 TSTPTFSVKDILNLEQ 21
RESULT 34
Q6UKJ7
ID Q6UKJ7 PRELIMINARY; PRT; 252 AA.
AC Q6UKJ7
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Timman (Fragment).
GN Name=tim;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=F-775F, S-438S, S-968F, S-565F, S-483F, US-255F, S-26F, S-510S,
RC S-521S, S-549S, and P-96S;
RX PubMed=15126403; DOI=10.1534/genetics.166.4.1845;
RA Balakirev E.S., Ayala F.J.;
RT "Nucleotide Variation in the tinman and bagpipe Homeobox Genes of
RT Drosophila melanogaster."
RL Genetics 166:1845-1856(2004).
DR EMBL; AY368098; AAQ75408.1; -
DR EMBL; AY368099; AAQ75409.1; -
DR EMBL; AY368100; AAQ75410.1; -
DR EMBL; AY368101; AAQ75411.1; -
DR EMBL; AY36


```
DR EMBL; AY368084; AAQ75394.1; --
FT NON_TER 252 252
SQ SEQUENCE 252 AA; 27337 MW; B41CC3F6C9EAB1E6 CRC64;

Query Match 53.9%; Score 55; DB 2; Length 252;
Best Local Similarity 64.7%; Pred. No. 0.89;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 TPFSVKDILSLPEQXRA 20
   |||||:::|
Db 36 TPFSVKDILNMVNQTEA 52

RESULT 37
Q6UKL4 PRELIMINARY; PRT; 252 AA.
AC Q6UKL4;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Tinman (Fragment).
GN Name=tin;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S-5F;
RX PubMed=15126403; DOI=10.1534/genetics.166.4.1845;
RA Balakirev E.S., Ayala F.J.;
RT "Nucleotide Variation in the tinman and bagpipe Homeobox Genes of
   Drosophila melanogaster.";
RL Genetics 166:1845-1856(2004).
DR EMBL; AY368081; AAQ75391.1; --
FT NON_TER 252 252
SQ SEQUENCE 252 AA; 27257 MW; DD563D824C61DC42 CRC64;

Query Match 53.9%; Score 55; DB 2; Length 252;
Best Local Similarity 64.7%; Pred. No. 0.89;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 TPFSVKDILSLPEQXRA 20
   |||||:::|
Db 36 TPFSVKDILNMVNQTEA 52

RESULT 38
Q6UKL5 PRELIMINARY; PRT; 252 AA.
AC Q6UKL5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Tinman (Fragment).
GN Name=tin;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S-258S;
RX PubMed=15126403; DOI=10.1534/genetics.166.4.1845;
RA Balakirev E.S., Ayala F.J.;
RT "Nucleotide Variation in the tinman and bagpipe Homeobox Genes of
   Drosophila melanogaster.";
RL Genetics 166:1845-1856(2004).
DR EMBL; AY368080; AAQ75390.1; --
DR EMBL; AY368082; AAQ75392.1; --
```

```
DR EMBL; AY368083; AAQ75393.1; --
DR EMBL; AY368085; AAQ75395.1; --
DR EMBL; AY368086; AAQ75396.1; --
DR EMBL; AY368088; AAQ75398.1; --
DR EMBL; AY368089; AAQ75399.1; --
DR EMBL; AY368090; AAQ75400.1; --
DR EMBL; AY368091; AAQ75401.1; --
DR EMBL; AY368092; AAQ75402.1; --
DR EMBL; AY368093; AAQ75403.1; --
DR EMBL; AY368094; AAQ75404.1; --
DR EMBL; AY368095; AAQ75405.1; --
DR EMBL; AY368096; AAQ75406.1; --
DR EMBL; AY368079; AAQ75389.1; --
FT NON_TER 252 252
SQ SEQUENCE 252 AA; 27306 MW; 12B1752DAB31DC50 CRC64;

Query Match 53.9%; Score 55; DB 2; Length 252;
Best Local Similarity 64.7%; Pred. No. 0.89;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 TPFSVKDILSLPEQXRA 20
   |||||:::|
Db 36 TPFSVKDILNMVNQTEA 52

RESULT 39
Q6UKL7 PRELIMINARY; PRT; 252 AA.
AC Q6UKL7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Tinman (Fragment).
GN Name=tin;
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=10F;
RX PubMed=15126403; DOI=10.1534/genetics.166.4.1845;
RA Balakirev E.S., Ayala F.J.;
RT "Nucleotide Variation in the tinman and bagpipe Homeobox Genes of
   Drosophila melanogaster.";
RL Genetics 166:1845-1856(2004).
DR EMBL; AY368078; AAQ75388.1; --
FT NON_TER 252 252
SQ SEQUENCE 252 AA; 27350 MW; EFS474CA6EAD68A1 CRC64;

Query Match 53.9%; Score 55; DB 2; Length 252;
Best Local Similarity 64.7%; Pred. No. 0.89;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 TPFSVKDILSLPEQXRA 20
   |||||:::|
Db 36 TPFSVKDILNMVNQTEA 52

RESULT 40
Q6UKL8 PRELIMINARY; PRT; 252 AA.
AC Q6UKL8;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Tinman (Fragment).
GN Name=tin;
OS Drosophila sechellia (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
```



```

OX NCBI_TaxID=7238;
RN [1]
RP SEQUENCE FROM N.A.
EX PubMed=15126403; DOI=10.1534/genetics.166.4.1845;
RA Balakirev E.S., Ayala F.J.;
RT "Nucleotide Variation in the tinman and bagpipe Homeobox Genes of
RL Drosophila melanogaster.";
RL Genetics 166:1845-1856(2004).
DR EMBL; AY368077; AAQ75387.1; -.
FT NON_TER 252
SQ SEQUENCE 252 AA; 27404 MW; 29FA605AD90A82E3 CRC64;

Query Match 53.9%; Score 55; DB 2; Length 252;
Best Local Similarity 64.7%; Pred. No. 0.89;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 TPFSVKDILSLPEQXRA 20
DB 36 TPFSVKDILNMVNQTEA 52

```

Search completed: April 13, 2005, 17:16:16
Job time : 24.1453 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2005, 16:58:19 ; Search time 269,966 Seconds
(without alignments)
375.348 Million cell updates/sec

Title: us-09-998-861-13

Perfect score: 1396

Sequence: 1 MEGNLQTFLLNTQLAQS.....LLLRKHRAAFSVLSLGAHSV 262

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_16Dec04.*

- 1: geneseqp1980s.*
- 2: geneseqp1990s.*
- 3: geneseqp2000s.*
- 4: geneseqp2001s.*
- 5: geneseqp2002s.*
- 6: geneseqp2003as.*
- 7: geneseqp2003bs.*
- 8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1396	100.0	262	5	ABP53047 Human Gro
2	590	42.3	277	5	ABB10102 Human hom
3	590	42.3	277	8	ADP03425 Human Gtx
4	558.5	40.0	365	8	ADG09852 Mouse NK-
5	557	38.9	367	5	ABB10101 Human hom
6	557	38.9	367	5	ADG09892 Human NKX
7	557	39.9	367	8	ADG09851 Human NK-
8	347	24.9	610	4	ABE711317 Drosophil
9	301	21.6	60	7	ABR82699 Nkx6.1 ho
10	301	21.6	60	8	ADP03435 Human Nkx
11	297	21.3	60	8	ADP03429 Mouse Gtx
12	297	21.3	60	8	ADP03427 Human Gtx
13	219.5	15.7	322	4	ABE60043 Drosophil
14	219.5	15.7	327	4	ABE66095 Drosophil
15	218	15.6	333	2	AAW86019 Murine ho
16	217	15.5	284	3	AAE26767 Rat insul
17	217	15.5	284	4	ABP35426 Secretory
18	217	15.5	284	6	ABP55166 Mouse tra
19	217	15.5	284	7	ADG09827 Mouse pdx
20	217	15.5	284	8	ADG09827 Mouse pdx
21	217	15.5	284	8	ADQ09818 Mouse pan
22	217	15.5	284	8	ADQ09826 Mouse pan
23	217	15.5	284	8	ADQ09826 Mouse pan
24	217	15.5	284	8	ADQ09825 Mouse pan
25	214.5	15.4	838	4	ABB58132 Drosophil

26	212.5	15.2	328	6	ADA83730	AdA83730 Human DLX
27	212.5	15.2	330	2	AAW86020	AAW86020 Human hom
28	211	15.1	284	2	AAW83396	AAW83396 Mouse ins
29	211	15.1	284	5	AAU99330	AAU99330 Mouse ins
30	211	15.1	284	5	AAU99323	AAU99323 Mouse ins
31	210.5	15.1	544	4	ABB61602	ABB61602 Drosophil
32	209.5	15.0	722	4	ABB63899	ABB63899 Drosophil
33	209	15.0	263	4	ABB61848	ABB61848 Drosophil
34	208	14.9	300	3	RAY84616	AY84616 A chicken
35	206.5	14.8	217	5	ABG96337	ABG96337 Human ova
36	206	14.8	230	3	RAY58610	AY58610 Protein r
37	206	14.8	291	8	ADQ17197	ADQ17197 Human sof
38	205.5	14.7	217	4	AAE10922	AAE10922 Human HOX
39	205.5	14.7	217	8	ADJ64361	ADJ64361 Cartilage
40	205.5	14.7	373	7	ADJ69169	ADJ69169 Human hea
41	204	14.6	166	8	ADL18317	ADL18317 Westcon mo
42	204	14.6	310	7	ADC31583	ADC31583 Human nov
43	202.5	14.5	306	4	ABB63596	ABB63596 Drosophil
44	202	14.5	283	8	ADQ09821	ADQ09821 Hamster p
45	202	14.5	283	8	ADQ09829	ADQ09829 Hamster p

ALIGNMENTS

RESULT 1
ABP53047
ID ABP53047 standard; protein; 262 AA.
XX
AC ABP53047;
XX
DT 06-NOV-2002 (first entry)
XX
DE Human Groucho-interacting protein Nkx6.3 SEQ ID NO:13.
XX
KW Human; Groucho-interacting protein; GIP; Nkx; Nkx6.3; differentiation;
KW Groucho corepressor protein.
XX
OS Homo sapiens.
XX
PN WO200242413-A2.
XX
PD 30-MAY-2002.
XX
PF 01-NOV-2001; 2001WO-IB002835.
XX
PR 01-NOV-2000; 2000US-0245002P.
XX
PA (NEUR-) NEURONOVA AB.
XX
PI Ericson J;
XX
DR WPI, 2002-657380/70.
XX
DR N-PSDB; ABQ75432.
XX
PT Guiding fate of differentiation of cell into cell type by contacting cell
PT with Groucho-interacting protein to form a complex which represses DNA
PT transcription and suppresses alternative pathways of differentiation.
XX
PS Claim 25; Page 89-90; 116pp; English.
XX
CC The present invention describes a method (M1) for guiding the fate of
CC differentiation of a cell into a specific cell type, by providing a
CC sample comprising the cell, contacting the sample with a Groucho-
CC interacting protein (GIP) to result in the formation of a complex between
CC GIP and a Groucho corepressor protein (I), where the complex represses
CC DNA transcription and suppresses alternative pathways of differentiation.
CC A purified complex (PCI), comprising a first polypeptide comprising a GIP
CC and the second polypeptide comprising a Groucho corepressor protein, can
CC be used for identifying an agent which modulates the stability or
CC activity of the complex, or that disrupts a polypeptide complex. PCI is
CC useful for screening the candidate substance interacting with the complex
CC or with (I); for inhibiting the guided differentiation of a cell

CC resulting in the impairment of ventral patterning; for identifying a
 CC polypeptide complex in a subject; and determining altered expression of a
 CC polypeptide in a subject. M1 is useful for guiding the fate of
 CC differentiation of a cell into a specific cell type. The present sequence
 CC represents a human GIP designated Nkx6.3, which is used in the
 CC exemplification of the present invention

XX
 SQ Sequence 262 AA;

Query Match 100.0%; Score 1396; DB 5; Length 262;
 Best Local Similarity 100.0%; Pred. No. 3e-131;
 Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MESNLQGTFLNNTLAQFSEMKAPMCOYVQNSFYKLSPPGLGQLAAGTGHGTTDLS 60
 DB 1 MESNLQGTFLNNTLAQFSEMKAPMCOYVQNSFYKLSPPGLGQLAAGTGHGTTDLS 60
 QY 61 RVATPNSLLSGYPHVAGFGLSSQGVYGPVQGSFSGKAGNEYTRTRNCWADTGDWR 120
 DB 61 RVATPNSLLSGYPHVAGFGLSSQGVYGPVQGSFSGKAGNEYTRTRNCWADTGDWR 120
 QY 121 GSAPCGNTPDPLSDTIHKKHTRPTFTGHQIFALEKTFEQTXYLAGPERARLAYSLGWT 180
 DB 121 GSAPCGNTPDPLSDTIHKKHTRPTFTGHQIFALEKTFEQTXYLAGPERARLAYSLGWT 180
 QY 181 ESQKVFQNRRTKWRKKSALPSSSTPRAPGASGDRAASENDEYNKPLDPSDDEK 240
 DB 181 ESQKVFQNRRTKWRKKSALPSSSTPRAPGASGDRAASENDEYNKPLDPSDDEK 240
 QY 241 IRLLRKHRAAFSVLSLGAHSV 262
 DB 241 IRLLRKHRAAFSVLSLGAHSV 262

RESULT 2

ABBI0102
 ID ABBI0102 standard; protein; 277 AA.

AC ABBI0102;

DT 01-JUL-2002 (first entry)

DE Human homeobox protein Nkx6.2 amino acid sequence.

KW Human; homeobox protein Nkx6.1; homeobox protein Nkx6.2; ventral neuron;
 KW amyotrophic lateral sclerosis; spinal muscular atrophy;
 KW neurodegenerative disease; transcription factor.

OS Homo sapiens.

PN WO200218545-A1.

PD 07-MAR-2002.

PF 31-AUG-2001; 2001WO-US027256.

PR 01-SEP-2000; 2000US-00654462.

PA (UYCO) UNIV COLUMBIA NEW YORK.

PI Jessell TM, Briscoe J, Ericson J, Rubenstein JLR, Sander M;

DR WPI; 2002-329764/36.

DR N-PSDB; ABL56620.

XX Converting stem cell into ventral neuron useful for treating neural
 PT degeneration in a subject, by introducing nucleic acid or polypeptide
 PT expressing homeodomain transcription factor Nkx6.1 or Nkx6.2 protein.

PS Disclosure; Fig 19, 108pp; English.

XX The invention relates to a method for converting a stem cell (SC) into a
 CC ventral neuron, comprising introducing a nucleic acid or polypeptide

CC expressing homeodomain transcription factor Nkx6.1 or Nkx6.2 protein into
 CC a stem cell. Methods of the invention are useful for converting a stem
 CC cell into a ventral neuron, and for diagnosing a motor neuron
 CC degenerative disease such as amyotrophic lateral sclerosis or spinal
 CC muscular atrophy in a subject. Methods of the invention are also useful
 CC for treating neuronal degeneration in a subject. The method comprises
 CC implanting a neural stem cell in diseased neural tissue under conditions
 CC so that the stem cell is converted into motor neuron after implantation.
 CC The current sequence represents the human homeobox protein Nkx6.2 amino
 CC acid sequence

XX
 SQ Sequence 277 AA;

Query Match 42.3%; Score 590; DB 5; Length 277;
 Best Local Similarity 50.8%; Pred. No. 2.5e-50;
 Matches 133; Conservative 28; Mismatches 85; Indels 16; Gaps 8;

QY 1 MESNLQGTFLNNTQLA---QFSEMKAPMCOYVQNS--SFYKLSPPGLGQLAAGTGHGI 55
 DB 1 MDTNRPAGAFVLSAPLAALHNMAMKTSLFYALQGPAGKAPALGGLGAQLPLGTHGI 60
 QY 56 TDILSRPVATPNSLLSGYPHVAGFGLSSQGVYGPVQV---SFSKAGNEYTRTRNCW 112
 DB 61 SDILCRPVGAAGGGLGGLPRINGLA--SSAGVYFGPAAAVARGYPKPLAELPGRPIFW 118
 QY 113 ADT--GQWRGSRAPCGNTP-DPLSDTIHKKHTRPTFTGHQIFALEKTFEQTXYLAGPE 169
 DB 119 PGVVQGAPWR-DPRLAGPAPAGGVLDKDGKKKHSRPTFSGQIIFALEKTFEQTXYLAGPE 177
 QY 170 RARLAYSILGTMESQVYKVFQNRRTKWRKKSALPSSSTPRAPGASGDRAASEN--EDDE 227
 DB 178 RARLAYSILGTMESQVYKVFQNRRTKWRKKSALPSSSTPRAPGASGDRAASEN--EDDE 227
 QY 228 YNKPLDPSDDEKIRLLRKHR 249
 DB 238 YNRPDPSDDEKIRLLRKHK 259

RESULT 3

ADP03425

ID ADP03425 standard; protein; 277 AA.

AC ADP03425;

DT 29-JUL-2004 (first entry)

DE Human Gtx/Nkx6.2.

KW human; cytotoxic inhibitor; homeobox; Alzheimer's disease; Gtx/Nkx6.2.

OS Homo sapiens.

PN JP2004137207-A.

PD 13-MAY-2004.

PF 18-OCT-2002; 2002JP-00304194.

PR 18-OCT-2002; 2002JP-00304194.

PA (KEIO-) GH KEIO GIJUKU.

DR WPI; 2004-361574/34.

DR N-PSDB; ADP03424.

XX Novel cytotoxic inhibitor comprising a protein containing a homeobox e.g.
 PT human Gtx/Nkx6.2, or a vector expressing the protein, useful for treating
 PT Alzheimer's disease.

PS Example 6; SEQ ID NO 2; 57pp; Japanese.

XX The invention relates to a cytotoxic inhibitor comprising a protein
 CC containing a homeobox, or a vector expressing the protein containing the

XX DE Human homeobox protein Nkx6.1 amino acid sequence.
 XX KW Human; homeobox protein Nkx6.1; homeobox protein Nkx6.2; ventral neuron;
 KW neurodegenerative disease; transcription factor.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT Misc-difference 224
 FT /note= "histidine residue that is not encoded in either
 FT of the exon 1 or exon 2 containing DNA fragments (see
 FT ABL56817 and ABL56818)"
 XX WO200218545-A1.
 XX PD 07-MAR-2002.
 XX PF 31-AUG-2001; 2001WO-US027256.
 XX PR 01-SEP-2000; 2000US-00654462.
 XX PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX PI Jessell TM, Briscoe J, Ericson J, Rubenstein JLR, Sander M;
 XX WPI; 2002-329764/36.
 XX DR N-PSDB; ABL56817, ABL56818, ABL56819.
 XX PT Converting stem cell into ventral neuron useful for treating neural
 PT degeneration in a subject, by introducing nucleic acid or polypeptide
 PT expressing homeodomain transcription factor Nkx6.1 or Nkx6.2 protein.
 XX PS Disclosure; Fig 6; 108pp; English.
 XX CC The invention relates to a method for converting a stem cell (SC) into a
 CC ventral neuron, comprising introducing a nucleic acid or polypeptide
 CC expressing homeodomain transcription factor Nkx6.1 or Nkx6.2 protein into
 CC a stem cell. Methods of the invention are useful for converting a stem
 CC cell into a ventral neuron, and for diagnosing a motor neuron
 CC degenerative disease such as amyotrophic lateral sclerosis or spinal
 CC muscular atrophy in a subject. Methods of the invention are also useful
 CC for treating neuronal degeneration in a subject. The method comprises
 CC implanting a neural stem cell in diseased neural tissue under conditions
 CC so that the stem cell is converted into motor neuron after implantation.
 CC The current sequence represents the human homeobox protein Nkx6.1 amino
 CC acid sequence
 XX SQ Sequence 367 AA;
 Query Match 39.9%; Score 557; DB 5; Length 367;
 Best Local Similarity 41.7%; Pred. No. 7.7e-47;
 Matches 145; Conservative 27; Mismatches 74; Indels 102; Gaps 13;
 QY 1 MESNQGTFLLNNTOLA---QFSENKAPM-----CQYSVNSFYKLSPP--- 41
 DB 7 MEGTRQSAFLSSSPPLAALHNSAEMKTLTPAAYVPLPAGPPSSSSSSSSPPLGT 66
 QY 42 ----GLGP-----QLAAGTGHGTTDILSRP-----VATPNS----- 69
 DB 67 HNPGLKPPATGGLSSLPQQLSAATPHGINILSRPMPVAGALPASPSGSSS 126
 QY 70 SSSSAS 93
 DB 127 SSSSAS 183
 QY 94 ---VGSFSGKAGNEVTRTENCWADTGOD--WEGSARPCGNTDPD---LSDTIHKKKTRP 145
 DB 184 VAAVGRYKPLAEPLGRTPIFWGVGMQSPWRDLAC--TPHOGSILLDDKGRKTRP 241
 QY 146 TFTGHQIFALEKTEFQTKYLAGEPARLAYSLGTMESQVKVWFQNRRTKWKHAAEMAT 205
 DB 302 AKKKQDSETERLKGAENEEDDDYDKPLDPSNDEKIKITQLLKKHKSS 349
 XX DE Human homeobox protein Nkx6.1 amino acid sequence.
 XX KW Human; homeobox protein Nkx6.1; homeobox protein Nkx6.2; ventral neuron;
 KW neurodegenerative disease; transcription factor.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT Misc-difference 224
 FT /note= "histidine residue that is not encoded in either
 FT of the exon 1 or exon 2 containing DNA fragments (see
 FT ABL56817 and ABL56818)"
 XX WO200218545-A1.
 XX PD 07-MAR-2002.
 XX PF 31-AUG-2001; 2001WO-US027256.
 XX PR 01-SEP-2000; 2000US-00654462.
 XX PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX PI Jessell TM, Briscoe J, Ericson J, Rubenstein JLR, Sander M;
 XX WPI; 2002-329764/36.
 XX DR N-PSDB; ABL56817, ABL56818, ABL56819.
 XX PT Converting stem cell into ventral neuron useful for treating neural
 PT degeneration in a subject, by introducing nucleic acid or polypeptide
 PT expressing homeodomain transcription factor Nkx6.1 or Nkx6.2 protein.
 XX PS Disclosure; Fig 6; 108pp; English.
 XX CC The invention relates to a method for converting a stem cell (SC) into a
 CC ventral neuron, comprising introducing a nucleic acid or polypeptide
 CC expressing homeodomain transcription factor Nkx6.1 or Nkx6.2 protein into
 CC a stem cell. Methods of the invention are useful for converting a stem
 CC cell into a ventral neuron, and for diagnosing a motor neuron
 CC degenerative disease such as amyotrophic lateral sclerosis or spinal
 CC muscular atrophy in a subject. Methods of the invention are also useful
 CC for treating neuronal degeneration in a subject. The method comprises
 CC implanting a neural stem cell in diseased neural tissue under conditions
 CC so that the stem cell is converted into motor neuron after implantation.
 CC The current sequence represents the human homeobox protein Nkx6.1 amino
 CC acid sequence
 XX SQ Sequence 367 AA;
 Query Match 39.9%; Score 557; DB 5; Length 367;
 Best Local Similarity 41.7%; Pred. No. 7.7e-47;
 Matches 145; Conservative 27; Mismatches 74; Indels 102; Gaps 13;
 QY 1 MESNQGTFLLNNTOLA---QFSENKAPM-----CQYSVNSFYKLSPP--- 41
 DB 7 MEGTRQSAFLSSSPPLAALHNSAEMKTLTPAAYVPLPAGPPSSSSSSSSPPLGT 66
 QY 42 ----GLGP-----QLAAGTGHGTTDILSRP-----VATPNS----- 69
 DB 67 HNPGLKPPATGGLSSLPQQLSAATPHGINILSRPMPVAGALPASPSGSSS 126
 QY 70 SSSSAS 93
 DB 127 SSSSAS 183
 QY 94 ---VGSFSGKAGNEVTRTENCWADTGOD--WEGSARPCGNTDPD---LSDTIHKKKTRP 145
 DB 184 VAAVGRYKPLAEPLGRTPIFWGVGMQSPWRDLAC--TPHOGSILLDDKGRKTRP 241
 QY 146 TFTGHQIFALEKTEFQTKYLAGEPARLAYSLGTMESQVKVWFQNRRTKWKHAAEMAT 205
 DB 302 AKKKQDSETERLKGAENEEDDDYDKPLDPSNDEKIKITQLLKKHKSS 349
 XX DE Human homeobox protein Nkx6.1 amino acid sequence.
 XX KW Human; homeobox protein Nkx6.1; homeobox protein Nkx6.2; ventral neuron;
 KW neurodegenerative disease; transcription factor.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT Misc-difference 224
 FT /note= "histidine residue that is not encoded in either
 FT of the exon 1 or exon 2 containing DNA fragments (see
 FT ABL56817 and ABL56818)"
 XX WO200218545-A1.
 XX PD 07-MAR-2002.
 XX PF 31-AUG-2001; 2001WO-US027256.
 XX PR 01-SEP-2000; 2000US-00654462.
 XX PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX PI Jessell TM, Briscoe J, Ericson J, Rubenstein JLR, Sander M;
 XX WPI; 2002-329764/36.
 XX DR N-PSDB; ABL56817, ABL56818, ABL56819.
 XX PT Converting stem cell into ventral neuron useful for treating neural
 PT degeneration in a subject, by introducing nucleic acid or polypeptide
 PT expressing homeodomain transcription factor Nkx6.1 or Nkx6.2 protein.
 XX PS Disclosure; Fig 6; 108pp; English.
 XX CC The invention relates to a method for converting a stem cell (SC) into a
 CC ventral neuron, comprising introducing a nucleic acid or polypeptide
 CC expressing homeodomain transcription factor Nkx6.1 or Nkx6.2 protein into
 CC a stem cell. Methods of the invention are useful for converting a stem
 CC cell into a ventral neuron, and for diagnosing a motor neuron
 CC degenerative disease such as amyotrophic lateral sclerosis or spinal
 CC muscular atrophy in a subject. Methods of the invention are also useful
 CC for treating neuronal degeneration in a subject. The method comprises
 CC implanting a neural stem cell in diseased neural tissue under conditions
 CC so that the stem cell is converted into motor neuron after implantation.
 CC The current sequence represents the human homeobox protein Nkx6.1 amino
 CC acid sequence
 XX SQ Sequence 367 AA;
 Query Match 39.9%; Score 557; DB 5; Length 367;
 Best Local Similarity 41.7%; Pred. No. 7.7e-47;
 Matches 145; Conservative 27; Mismatches 74; Indels 102; Gaps 13;


```

QY      1 MESNLQGTFLLNNTOLA---OFSENKAPM-----COYSVQNSEFYKLSP--   41
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      7 MEGTROSAPFLLSPPPLAALHGWAEKMTPLYPAAYPPLPAGPPSSSSSSSSSSPSPLGT 66
                                     :|::|
QY     42 ---GLGP-----QLAAGTPHGITDILSRP-----VATPNSS-----   69
                                     :|::|
Db     67 HNPGLLKPPATGCLSSLGSPPOOLSAATPHGINLISRPMPVNSGAALPSAGSGSSSS 126
                                     :|::|
QY     70 -----LLSGYPHVAGFGGLS-----SGGVTYGQP-----   93
                                     :|::|
Db    127 SSSASASSASAAAAAAAASPPAGLLAGLPR---FSSLSPPPPGLYFSPSAAA 183
                                     :|::|
QY     94 ---VGSFSKAGNEYTRTNCHWADTGOD--WRGSARPCGNTPDP---LSTTIHKKHTRP 145
                                     :|::|
Db    184 VAAVGRYPKPLAELFCRTPIFPWGVMQSPPPWRDLAC--TPHQGSTLLDKDGKRKHTRP 241
                                     :|::|
QY    146 TPTGHQIFALEKTFEQTKYLACPEARLAYSLGHTESQVKWFONRRKTWKKSALPESS 205
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    242 TFSGGQIFALEKTFEQTKYLACPEARLAYSLGHTESQVKWFONRRKTWKKHAEMAT 301
                                     :|::|
QY    206 STTRAPGGASGRASAEN--EDEYNKPLDPDSDEKIRLLLRKHRAA 251
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    302 AKKKODSETERUKGASENEEDDDYNNKPLDNSDDEKITOLLKKHKS 349

```

RESULT 7	
ADQ09851	
ID	ADQ09851 standard; protein; 367 AA.
XX	
AC	ADQ09851;
XX	
DT	23-SEP-2004 (first entry)
XX	
DE	Human NK-2 class homeobox protein, NKx6.1.
XX	
XX	Human; islet cell differentiation transcription factor;
KW	insulin-dependent diabetes; insulin; somatic cell;
KW	insulin-producing cell; islet cell; NeuroD; ngn3; neurogenin 3; Pax6;
KW	paired-box transcription factor 6; Pax4;
KW	paired-box transcription factor 4; Nkx2.2; NK-2 class homeobox protein;
KW	Nkx6.1; Isl-1; Islet factor 1; Pdx-1;
KW	pancreatic and duodenal homeobox protein; BCT; betacellulin.
XX	
OS	Homo sapiens.
XX	
PN	US2004132679-A1.
XX	
PD	08-JUL-2004.
XX	
PF	03-SEP-2003; 2003US-00654102.
XX	
PR	03-SEP-2002; 2002US-0407743P.
XX	
PA	(BAYU) BAYLOR COLLEGE MEDICINE.

XX Use of an islet cell differentiation transcription factor polypeptide or
 PT its homologue or analog for treating a mammal for insulin-dependent
 PT diabetes, increasing an insulin level in a somatic cell, or generating an
 PT insulin-producing cell.
 XX
 XX Claim 109; SEQ ID NO 89; 190pp; English.
 PS
 XX
 CC The invention relates to the use of an islet cell differentiation
 CC transcription factor polypeptide or its homologue or analogue for
 CC treating a mammal for insulin-dependent diabetes, increasing an insulin
 CC level in a somatic cell, or generating an insulin-producing cell. Also
 CC included are a method of treating a mammal for insulin-dependent

diabetes, a method of increasing an insulin level in a somatic cell, a method of generating an insulin-producing cell, a therapeutic composition comprising an isolated islet cell differentiation transcription factor polypeptide (and/or an isolated nucleic acid expressing the polypeptide), an insulin-producing cell comprising a vector (where the vector comprises a nucleic acid sequence encoding an islet cell differentiation transcription factor), an insulin-producing cell (generated by a method comprising obtaining a somatic cell and transfecting the cell with a vector comprising a nucleic acid sequence encoding an islet cell differentiation transcription factor, where in the transfecting step the cell produces insulin), a method of generating at least one pancreatic islet, and a composition (comprising: NeuroD or ngn3 polypeptide or a polynucleotide expressing a NeuroD or ngn3 polypeptide and betacellulin polypeptide or a polynucleotide expressing a betacellulin polypeptide). The islet cell differentiation transcription factor polypeptide is NeuroB, ngn3 (neurogenin 3), Pax6 (paired-box transcription factor 6), Pax4 (paired-box transcription factor 4), Nkx2.2 (NK-2 class homeobox protein), Nkx6.1, Isl-1 (Islet factor 1), Pdx-1 (pancreatic and duodenal homeobox protein), BCT (betacellulin) or their combinations. The islet cell differentiation transcription factor polypeptide or its homologue or analogue is useful for treating a mammal for insulin-dependent diabetes, increasing an insulin level in a somatic cell, or generating an insulin-producing cell. The present sequence is an islet cell differentiation transcription factor polypeptide as detailed above.

```

SQ      Sequence 367 AA;

Query Match          39.9%; Score 557; DB 8; Length 367;
Best Local Similarity 41.7%; Pred. No. 7.7e-47;
Matches 145; Conservative 27; Mismatches 74; Indels 102; Gaps 13;

QY      1 MESNLOCTILLNNTOLA---QFSEMKAPM-----CQYSVQNSFYKLSPP--- 41
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       7 MEGTROSAFLLSPPLAALHLSMAEMKTPIYPAAYPELPAGPFPSSSSSSSSPSPLPTG 66
                                     :|::|::|::|::|::|::|::|::|::|::|
QY      42 ----GLGP-----QLAAGTGHGTTDILSRP-----VATPNSS----- 69
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       67 HNPGLKLKPATGCLSLGSPPOOLSNAATHGINNILSRPMPVASGAALPSAPSGSSSS 126
                                     :|::|::|::|::|::|::|::|::|::|::|
QY      70 -----LLGGYPHVAGFGGLS-----SQGVVYGPO--- 93
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     127 SSSSASASSASASASASASASASASASASAGLLAGLPR---FSSLSPPPPGLFYSPSAAA 183
                                     :|::|::|::|::|::|::|::|::|::|::|
QY      94 ----VGFSFKAGNEYPRTRNCWADTCQD--WRGSARPCGNTDPD---LSDTHKKKHTRP 145
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     184 VAAVGRYPRLAELPGRTPIFWGVNQSPWRDLAC--TPHQGSILLDKOGKRKHTRP 241
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY     146 TFTGHQIFALEKTFEOTKYLAGPERARLAYSLGMTESQVKWFQNRRTKRWKKSALPESS 205
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     242 TTSQQQIFALEKTFEOTKYLAGPERARLAYSLGMTESQVKWFQNRRTKRWKKAEMAT 301
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

QY     206 STPRAPGASGDRAASEN--EDDEYNKLPDDSDDEKIRLLLRKHRAA 251
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     302 AKKQODSETERLKGAENEIEEDDYNNKLPDPSDDDEKITOLLAKKHSS 349
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

RESULT 8	
ABB71317	Drosophila melanogaster polypeptide SEQ ID NO 40743 .
ID	Drosophila; developmental biology; cell signalling; insecticide;
XX	pharmaceutical.
XX	Drosophila melanogaster.
AC	XX
AC	XX
ABB71317;	WO200171042-A2 .
XX	XX
XX	XX
26-MAR-2002 (first entry)	
DT	
DT	
DE	
XX	
XX	
KW	
KW	
KW	
OS	
OS	
XX	
PN	
XX	
XX	


```

PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEXE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL15420.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 40743; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signaling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 610 AA;
SQ
Query Match 24.9%; Score 347; DB 4; Length 610;
Best Local Similarity 42.6%; Pred No. 1.9e-25;
Matches 87; Conservative 24; Mismatches 73; Indels 20; Gaps 6;
QY 49 ACTPHGIDILSR--PVATPNSLLSG-----YPHVAGFGGLS-----SQGVYXGPQVGS 96
Db 283 APNPHGIDITLSKPPPVTSAGLSALTGAGIPRFSIAAAAGWAQVLSQSGAPLTKTHAGH 342
QY 97 FSKAGNEYPTTRN-----CWADTGQDWGSRPCGNTPPDPLSDTIHKKHTRPTFTGH 150
Db 343 IVDRTHLVWPGLQGLVANPIAWRRLSNTMTSANGLSQSHQHPNSDKGKKHTRPTFTSGQ 402
QY 151 QIFALEKTFEQTLYLAGPERARLAYSLGWTESQVKVWFQNRRTKWRKKSALPSSSTPRA 210
Db 403 QIFALEKTFEQTLYLAGPERAKLAYALGWSQVKVWFQNRRTKWRKKEHAEMATAKEKQ 462
QY 211 PG-GASGDRAASENEDDEYNKPLD 233
Db 463 DDMGGDNDGDCSETWDSN-NESLD 485
RESULT 9
ABR82699
ID ABR82699 standard; peptide; 60 AA.
XX
XX ABR82699;
XX
XX 04-DEC-2003 (first entry)
XX
XX Nkx6.1 homeodomain protein fragment.
XX
XX Pluripotency determining factor; gp130; pluripotency; transgenic; PDF;
XX homeodomain.
XX
XX Unidentified.
XX
XX WO2003064463-A2.
XX
XX 07-AUG-2003.
XX
30-JAN-2003; 2003WO-GB000366.
30-JAN-2002; 2002GB-00002149.
(UYED-) UNIV EDINBURGH.
Chambers I, Smith AG;
WPI; 2003-646140/61.
New factor that acts intracellularly in the absence of gp130 activation
on a mouse ES cells or on stem cell from a non-permissive strain of mice,
useful for maintaining a cell in a pluripotent state.
Disclosure; Fig 3A; 126pp; English.
The invention relates to a factor that acts intracellularly and maintains
a cell in a pluripotent state in the absence of gp130 activation, or that
maintains or confers pluripotency of a human stem cell, on a mouse ES
cell and acts intracellularly in the absence of gp130 activation, or on a
stem cell from a non-permissive strain of mice. The factor has the
ability to stably proliferate pluripotent cells in culture as compared to
prior art. The factor, conjugate, nucleotide sequence and vector are
useful for maintaining a cell in a pluripotent state. A screening method
is provided for identifying a cDNA that confers a self-renewing
phenotype. The cells are useful for screening molecules that interfere
with the function of the factor. Sequences ABR82687-699 represent several
different classes of homeodomain protein, used in an alignment study with
a pluripotency determining factor (PDF) homeodomain fragment
Sequence 60 AA;
SQ
Query Match 21.6%; Score 301; DB 7; Length 60;
Best Local Similarity 94.9%; Pred No. 2.7e-22;
Matches 56; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 140 KKHTRPTFTGHQIFALEKTFEQTLYLAGPERARLAYSLGWTESQVKVWFQNRRTKWRKK 198
Db 1 RKHTRPTFTSGQIFALEKTFEQTLYLAGPERARLAYSLGWTESQVKVWFQNRRTKWRKK 59
RESULT 10
ADP03435
ID ADP03435 standard; protein; 60 AA.
XX
XX AC ADP03435;
XX
XX DT 29-JUL-2004 (first entry)
XX
XX DE Human Nkx6.1homeobox.
XX
XX KW human; cytotoxic inhibitor; homeobox; Alzheimer's disease; Nkx6.1.
XX
XX OS Homo sapiens.
XX
XX PN JP2004137207-A.
XX
XX PD 13-MAY-2004.
XX
XX PF 18-OCT-2002; 2002JP-00304194.
XX
XX PR 18-OCT-2002; 2002JP-00304194.
XX
XX PA (KEIO-) GH KEIO GIJUKU.
XX
XX DR WPI; 2004-361574/34.
XX
XX DR N-PSDB; ADP03434.
XX
XX Novel cytotoxic inhibitor comprising a protein containing a homeobox e.g.
XX human Gtx/Nkx6.2, or a vector expressing the protein, useful for treating
XX Alzheimer's disease.
XX
XX Example 6; SEQ ID NO 12; 57pp; Japanese.

```


XX The invention relates to a cytotoxic inhibitor comprising a protein
 CC containing a homeobox, or a vector expressing the protein containing the
 CC homeobox. The method is useful for inhibiting cytotoxicity associated
 CC with Alzheimer's disease. A pharmaceutical composition is useful for
 CC treatment or prevention of disease such as Alzheimer's disease associated
 CC with cytotoxicity. A cytotoxic inhibitor is useful for inhibiting
 CC cytotoxicity associated with Alzheimer's disease. The cytotoxic inhibitor
 CC is useful for treating Alzheimer's disease. The cytotoxic inhibitor
 CC effectively inhibits the cytotoxicity. The present sequence represents
 CC the amino acid sequence of human Nkx6.1 homeobox.
 XX
 SQ Sequence 60 AA;

Query Match 21.6%; Score 301; DB 8; Length 60;
 Best Local Similarity 94.9%; Pred. No. 2.7e-22;
 Matches 56; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 140 KKHTPTTGHQIFALEKTFEQTKYLAGEPARLAYSLGTMESQVKVWFQNRRTKWRKK 198
 Db 1 RKHTRPTSGQOIFALEKTFEQTKYLAGEPARLAYSLGTMESQVKVWFQNRRTKWRKK 59

RESULT 11
 ADP03429
 ID ADP03429 standard; protein; 60 AA.

XX AC ADP03429;
 XX DT 29-JUL-2004 (first entry)
 XX DE Mouse Gtx/Nkx6.2 homeobox.
 XX KW mouse; cytotoxic inhibitor; homeobox; Alzheimer's disease; Gtx/Nkx6.2.
 XX OS Mus musculus.
 XX PN JP2004137207-A.
 XX PD 13-MAY-2004.
 XX PF 18-OCT-2002; 2002JP-00304194.
 XX PR 18-OCT-2002; 2002JP-00304194.
 XX PA (KEIO-) GH KEIO GIJUKU.
 XX DR WPI; 2004-361574/34.
 XX DR N-PSDB; ADP03428.

XX Novel cytotoxic inhibitor comprising a protein containing a homeobox e.g.
 PT human Gtx/Nkx6.2, or a vector expressing the protein, useful for treating
 PT Alzheimer's disease.

XX Example 6; SEQ ID NO 6; 57pp; Japanese.

XX The invention relates to a cytotoxic inhibitor comprising a protein
 CC containing a homeobox, or a vector expressing the protein containing the
 CC homeobox. The method is useful for inhibiting cytotoxicity associated
 CC with Alzheimer's disease. A pharmaceutical composition is useful for
 CC treatment or prevention of disease such as Alzheimer's disease associated
 CC with cytotoxicity. A cytotoxic inhibitor is useful for inhibiting
 CC cytotoxicity associated with Alzheimer's disease. The cytotoxic inhibitor
 CC is useful for treating Alzheimer's disease. The cytotoxic inhibitor
 CC effectively inhibits the cytotoxicity. The present sequence represents
 CC the amino acid sequence of mouse Gtx/Nkx6.2 homeobox.
 XX

SQ Sequence 60 AA;

Query Match 21.3%; Score 297; DB 8; Length 60;
 Best Local Similarity 93.2%; Pred. No. 6.9e-22;
 Matches 55; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 140 KKHTPTTGHQIFALEKTFEQTKYLAGEPARLAYSLGTMESQVKVWFQNRRTKWRKK 198
 Db 1 KKHSPFTSGQOIFALEKTFEQTKYLAGEPARLAYSLGTMESQVKVWFQNRRTKWRKK 59

RESULT 12
 ADP03427
 ID ADP03427 standard; protein; 60 AA.

XX AC ADP03427;
 XX DT 29-JUL-2004 (first entry)
 XX DE Human Gtx/Nkx6.2 homeobox.
 XX KW human; cytotoxic inhibitor; homeobox; Alzheimer's disease; Gtx/Nkx6.2.
 XX OS Homo sapiens.
 XX PN JP2004137207-A.
 XX PD 13-MAY-2004.
 XX PF 18-OCT-2002; 2002JP-00304194.
 XX PR 18-OCT-2002; 2002JP-00304194.
 XX PA (KEIO-) GH KEIO GIJUKU.
 XX DR WPI; 2004-361574/34.
 XX DR N-PSDB; ADP03426.

XX Novel cytotoxic inhibitor comprising a protein containing a homeobox e.g.
 PT human Gtx/Nkx6.2, or a vector expressing the protein, useful for treating
 PT Alzheimer's disease.

XX Claim 2; SEQ ID NO 4; 57pp; Japanese.

XX The invention relates to a cytotoxic inhibitor comprising a protein
 CC containing a homeobox, or a vector expressing the protein containing the
 CC homeobox. The method is useful for inhibiting cytotoxicity associated
 CC with Alzheimer's disease. A pharmaceutical composition is useful for
 CC treatment or prevention of disease such as Alzheimer's disease associated
 CC with cytotoxicity. A cytotoxic inhibitor is useful for inhibiting
 CC cytotoxicity associated with Alzheimer's disease. The cytotoxic inhibitor
 CC is useful for treating Alzheimer's disease. The cytotoxic inhibitor
 CC effectively inhibits the cytotoxicity. The present sequence represents
 CC the amino acid sequence of human Gtx/Nkx6.2 homeobox.
 XX

SQ Sequence 60 AA;

Query Match 21.3%; Score 297; DB 8; Length 60;
 Best Local Similarity 93.2%; Pred. No. 6.9e-22;
 Matches 55; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 140 KKHTPTTGHQIFALEKTFEQTKYLAGEPARLAYSLGTMESQVKVWFQNRRTKWRKK 198
 Db 1 KKHSPFTSGQOIFALEKTFEQTKYLAGEPARLAYSLGTMESQVKVWFQNRRTKWRKK 59

RESULT 13
 ABB60043
 ID ABB60043 standard; protein; 322 AA.

XX AC ABB60043;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 6921.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical.

OS Drosophila melanogaster.
 XX WO200171042-A2.
 PN
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL04146.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 PS Disclosure; SEQ ID NO 6921; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 322 AA;
 Query Match 15.7%; Score 219.5; DB 4; Length 322;
 Best Local Similarity 34.6%; Pred. No. 4.8e-13;
 Matches 55; Conservative 29; Mismatches 50; Indels 25; Gaps 5;
 QY 78 AGFCGLSSQGVYGPQVGS-----FSKAGNEYPTTRNCWADTGODWRGSRAPCGNTP 130
 DB 43 AGYGIRSTYQHFGPGQGDGSPRPSRSGALGYPPPMHONS- --SGYHLGSYAPPCASPP 100
 QY 131 D----PLSDTIH-----KKKHTPTFTGHQIFALEKTFEOTKYLAGPERARLAYS 176
 DB 101 KDDPSISKCEDSLRVNGKGMKPKRTIYSSLOQLNRRFQRTQYLALPERAELAA 160
 QY 177 LGMTESQVKVWFONRTKWRK--KSALEPSSSTPRAPGG 213
 DB 161 LGLTQTQVKIWFQNRSSKYKMMKAAQGGTNSGMPLGG 199
 RESULT 14
 ABB66095
 ID ABB66095 standard; protein; 327 AA.
 XX
 AC ABB66095;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 25077.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 OS
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX

PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL10198.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 PS Disclosure; SEQ ID NO 25077; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 327 AA;
 Query Match 15.7%; Score 219.5; DB 4; Length 327;
 Best Local Similarity 34.6%; Pred. No. 4.9e-13;
 Matches 55; Conservative 29; Mismatches 50; Indels 25; Gaps 5;
 QY 78 AGFCGLSSQGVYGPQVGS-----FSKAGNEYPTTRNCWADTGODWRGSRAPCGNTP 130
 DB 43 AGYGIRSTYQHFGPGQGDGSPRPSRSGALGYPPPMHONS- --SGYHLGSYAPPCASPP 100
 QY 131 D----PLSDTIH-----KKKHTPTFTGHQIFALEKTFEOTKYLAGPERARLAYS 176
 DB 101 KDDPSISKCEDSLRVNGKGMKPKRTIYSSLOQLNRRFQRTQYLALPERAELAA 160
 QY 177 LGMTESQVKVWFONRTKWRK--KSALEPSSSTPRAPGG 213
 DB 161 LGLTQTQVKIWFQNRSSKYKMMKAAQGGTNSGMPLGG 199
 RESULT 15
 AAW86019
 ID AAW86019 standard; protein; 333 AA.
 XX
 AC AAW86019;
 XX
 DT 22-FEB-1999 (first entry)
 XX
 DE Murine homeobox HOX11 protein sequence.
 XX
 KW HOX11 gene; homeobox; transgenic; mouse; human; gene inactivation;
 KW aplenic; screening; developmental; pathological; thymus; T cell;
 KW organogenesis.
 XX
 OS Mus sp.
 XX
 PN US5850002-A.
 XX
 PD 15-DEC-1998.
 XX
 PF 13-SEP-1996; 96US-00712948.
 XX
 PR 20-APR-1994; 94US-00231728.
 XX
 PA (UNIW) UNIV WASHINGTON.

XX Kormeyer SJ;
 XX WPI; 1999-069841/06.
 XX Transgenic mouse with inactivated Hox11 gene - and transgenic mouse
 XX having human Hox11 gene, used for screening compounds.
 XX Disclosure; Col 15-18; 26pp; English.
 XX The invention provides a transgenic mouse homozygous for Hox11 gene
 CC inactivation. The Hox11 gene is inactivated by the insertion of a
 CC heterologous nucleic acid sequence into the Hox11 gene via homologous
 CC recombination, where the inactivation of the Hox11 gene prevents
 CC expression of the Hox11 gene, and whereby the mouse is asplenic. The
 CC mouse is used for screening for a compound, where the compound affects a
 CC developmental and/or pathological condition due to expression of the
 CC human Hox11 gene in the thymus or T cells that do not normally express
 CC the human Hox11 gene. Animals deficient in orphan homeobox gene function
 CC serve as models for organogenesis. The present sequence represents the
 CC murine homeobox Hox11 protein sequence
 XX Sequence 333 AA;
 Query Match 15.6%; Score 218; DB 2; Length 333;
 Best Local Similarity 29.9%; Pred. No. 7.1e-13;
 Matches 73; Conservative 29; Mismatches 78; Indels 64; Gaps 9;
 QY 36 YKLSPPGLPQLAAGTPHGTIDILSRPVATPNSLLSGYPHVGFGG-----LSSQ 86
 Db 74 YGAGGPG-GPGPGAGGGGACSMGLPGSYNNMALAGPGPGGGGGGAGGAGALSA 132
 QY 87 GYV-----YGPQVGSFSK-----AGNEYPTTRNCWADTGQD 119
 Db 133 GVIRVPAHRPLAGAVAHPOPLATGLTPVSPVAVGVNLTGLTFP-----WMSNRY 186
 QY 120 ---RGSARPCGTPDPLSDTHKKHTRPTTGHQIFALEKTFEQTLYAGPERARLAYS 176
 Db 187 TKDRFTGHPYQNRTPP-----KKKKPRTSFLQLICELEKRFHQRKYLASAERAAALAKA 240
 QY 177 LQMTESQVKVWFQNRRTKWRKKSALPSSSTPRAPGASGDRAASENEDDEYNK----- 230
 Db 241 LQMTDAQVKTWFQNRRTKWRRTABERAEARQQA-----NRILLQLQAEAFQKSLAQL 294
 QY 231 PLDP 234
 Db 295 PADP 298
 RESULT 16
 ID AAB26767
 XX AAB26767 standard; protein; 284 AA.
 XX AAB26767;
 XX 17-JAN-2001 (first entry)
 XX Rat insulin promoter factor 1 protein sequence.
 XX Amylin; production; secretory cell; blood glucose level regulation;
 KW diabetes mellitus; hypoglycaemia; osteoporosis; Paget's disease;
 KW hypercalcaemia; obesity; hypertension.
 XX Rattus sp.
 XX US6110707-A.
 XX 29-AUG-2000.
 XX 17-JAN-1997; 97US-00784582.
 XX 19-JAN-1996; 96US-00589028.
 PR 11-OCT-1996; 96US-0028279P;

XX (TEXA) UNIV TEXAS SYSTEM.
 PA (BETA-) BETAGENE INC.
 XX Newgard CB, Halban P, Normington KD, Thigpen AE, Quaade C;
 PI Kruse F, McGarry D, Clark SA;
 XX WPI; 2000-586352/55.
 DR N-PSDB; AAC55718.
 XX Producing mammalian amylin, useful for regulating blood glucose and
 PT insulin levels, e.g. for treating diabetes mellitus or hypoglycemia, by
 PT employing recombinantly engineered secretory cell lines.
 XX Disclosure; Col 149-152; 136pp; English.
 XX This invention relates to a method for producing mammalian amylin. The
 CC method relies on the use of a recombinantly engineered secretory cell
 CC line. The method comprises: (a) providing a starting secretory cell that
 CC has a regulated secretory pathway; (b) introducing, into the starting
 CC secretory cell, an amylin-encoding gene operatively linked to a first
 CC promoter; (c) selecting a secretory cell of (b) that exhibits increased
 CC production of biologically active amylin as compared to the starting
 CC secretory cell; and (d) culturing the selected secretory cell. Amylin is
 CC an insulin modulator, and the method results in antidiabetic, hypotensive
 CC and osteopathic activity. The amylin produced are useful for regulating
 CC blood glucose levels, as well as in modulating the circulating levels of
 CC insulin in a mammal. The amylin produced may be used in treating diabetes
 CC mellitus, hypoglycaemia, osteoporosis, Paget's disease, hypercalcaemia,
 CC obesity, hypertension, or any other disorder requiring amylin regulation.
 CC The invention includes cDNA and protein sequences (AAC55760 and AAC26771)
 CC representing human amylin. Sequences AAC55716-C55681 and AAC26765-B26777
 CC are used in examples of the method of the invention for the production of
 CC mammalian amylin
 XX Sequence 284 AA;
 Query Match 15.5%; Score 217; DB 3; Length 284;
 Best Local Similarity 30.3%; Pred. No. 7.1e-13;
 Matches 70; Conservative 26; Mismatches 95; Indels 40; Gaps 9;
 QY 25 PMCVSYQNSFVKLSPPGLGPGQ-----LAAGTPHGTIDILSRPVATPNSLLSGYPHV 77
 Db 33 PACLY-----MGRQPPPPPPPPFTSSLSLQSGSPDISPYEVPPLASDDDPAGAHHLHL 87
 QY 78 AGFGLSQGVYYPQVGSFSKAGNEYPTTR-----NCWADTGDWRGSARPC 126
 Db 88 PAQLGLAHPPP--GFPNGTEPGGLEENRVQLPFPWMKSTKAHAW--KGQ-WAGGA--- 139
 QY 127 GNTDPLSDTHKKHTRPTTGHQIFALEKTFEQTLYAGPERARLAYS LGMTESQVKV 186
 Db 140 -YTAEP-----EENKRTRTAYTRAQLLELEKEFLFNKYSRPRVELAVMLNLTERRHIKI 193
 QY 187 WFNRRRTKWRKKSALPSSSTPRAPGGA---SGRAASENEDDEYNKPLDP 234
 Db 194 WFNRRMKWKEEDKKRSSGTPSGGGGEEPEQDCAVTSGEELLAVPPLPP 244
 RESULT 17
 AAB35426
 ID AAB35426 standard; protein; 284 AA.
 XX AAB35426;
 XX 23-MAY-2001 (first entry)
 XX Secretory cell line protein production method protein SEQ ID NO: 6.
 XX Secretory cell line; protein production; diabetes; disease treatment;
 KW hormone.
 XX Unidentified.
 OS

PN US6194176-B1.
 PD 27-FEB-2001.
 PP 17-JAN-1997; 97US-00785271.
 XX 19-JAN-1996; 96US-00589028.
 XX (TEXA) UNIV TEXAS SYSTEM.
 PA (BETA-) BETAGENE INC.
 XX Newgard CB, Halban P, Normington KD, Clark SA, Thigpen AB;
 PI Quade C, Kruse F;
 XX WPI; 2001-217929/22.
 XX Producing human insulin involves transforming secretory host cell with
 PT exogenous polynucleotide comprising gene encoding insulin, and culturing
 PT host cell.
 XX Disclosure; Col 117-118; 107pp; English.
 XX The present invention describes a method of producing human insulin,
 CC involving transforming a secretory host cell with an exogenous insulin
 CC coding sequence and culturing the cell in conditions suitable for insulin
 CC expression. The invention also describes secretory cells useful as hosts.
 CC The method is also useful in the production of hormones and other
 CC proteins for use in disease treatment. The present sequence is a protein
 CC sequence used to demonstrate the method of the invention
 XX
 SQ Sequence 284 AA;
 Query Match 15.5%; Score 217; DB 4; Length 284;
 Best Local Similarity 30.3%; Pred. No. 7.1e-13;
 Matches 70; Conservative 26; Mismatches 95; Indels 40; Gaps 9;
 QY 25 PMCQSVQNSFYKLSPPGLGQ-----LAAGTGHGTTILSRPVATPNSLLSGYPHV 77
 DB 33 PACLY-----MGRQPPPPPPQFTSLGSLGQSPDPISPYEVPLASDDPAGAHLLHHL 87
 QY 78 AGFGGLSSQGVYQVGSFSGKAGNEYPTTR-----NCWADTGDWRGSRAPC 126
 DB 88 PAQLGLAHPPP--GFPNGTEPGGLEENRVLPFPWMKSTKAHAW--KGQ-WAGGA--- 139
 QY 127 GNTDPDLSDTIHKKKHTRPTFTGHQIFALEKTFEQTKYLAGPERARLAYSIGMTESQVKV 186
 DB 140 -YTAEP-----EENKTRTAYTRAQILLEKEFLFNKYSRPRVELAVMLNLTHERHIKI 193
 QY 187 WFNQRRTKWRKKSALPSSSTPRAPGGA---SGDRAASENEDDEYNKPLDP 234
 DB 194 WFNQRRMKWKEEDKKRSSGTSPGGGGEEPEQDCAVTSGBELLAVPPLPP 244
 RESULT 18
 ID ABP55166 standard; protein; 284 AA.
 XX
 AC ABP55166;
 XX
 DT 11-FEB-2003 (first entry)
 XX
 DE Mouse transcription factor Pdx1.
 XX
 KW Mouse; transcription factor; Pdx1; stem cell; differentiation; beta-cell;
 KW insulin; diabetes; hyperglycaemia; glucose intolerance; antidiabetic;
 KW hypoglycaemic; gene therapy.
 XX
 OS Mus musculus.
 XX
 PN WO200286107-A2.
 XX
 XX 31-OCT-2002.
 DT

PF 19-APR-2002; 2002WO-EP004362.
 XX
 PR 19-APR-2001; 2001US-0284531P.
 XX
 PA (DEVE-) DEVELOPENTWICKLUNGSBIOLOGISCHE FORSCH.
 PA (PFLA-) INST PFLANZENGENETIK & KULTURPFLANZENFOR.
 XX
 PI Wobus AM, St-Onge L, Blyszczuk P, Hoffmann U;
 XX WPI; 2003-075629/07.
 DR N-PSDB; ABV75967.
 XX
 XX Differentiating stem cells into insulin-producing cells useful for
 PT treating pancreatic diseases, by culturing stem cells in suitable medium
 PT and activating gene involved in beta-cell differentiation.
 XX
 PS Disclosure; Page 48-49; 62pp; English.
 XX
 CC The present sequence is the protein sequence of the murine homeobox-
 CC containing transcription factor Pdx1 (also referred to Idx1, STFI and
 CC IPFI). Pdx1 is expressed in all cells of the pancreatic buds during
 CC development but is restricted to the beta-cells in adult animals. The
 CC invention provides a claimed method for differentiating stem cells
 CC (especially embryonic, adult or somatic stem cells and primordial germ
 CC cells) into insulin-producing cells. This involves culturing stem cells
 CC in a suitable medium and activating at least one gene involved in beta-
 CC cell differentiation. Preferred genes including Pdx1, Pax4, Pax6 and ngn3
 CC (see ABV75967-70). Gene activation comprises the delivery of the gene
 CC into stem cells using a viral delivery system, or the delivery of a
 CC protein product of the gene into stem cells. The insulin-producing cells
 CC can be transplanted into animals or human for treatment of pancreatic
 CC diseases, metabolic syndrome and metabolic disorders with impaired
 CC glucose levels such as diabetes, hyperglycaemia and impaired glucose
 CC tolerance (Claimed). The cells can also be used to identify compounds
 CC which stimulate beta-cell differentiation, insulin secretion or glucose
 CC responsiveness. Differentiated beta-cells can be used to study the toxic
 CC and other effects of exogenous compounds on beta-cell function. In an
 CC example from the invention, Pdx1 cDNA was inserted into expression vector
 CC pACMV.plpA under the control of the cytomegalovirus promoter. The
 CC transgene was introduced into mouse R1 embryonic stem (ES) cells by
 CC electroporation. These ES cells were cultivated as embryoid bodies. At
 CC day 15 of plating, 22% of the Pdx1+ ES cells were positive for insulin
 XX
 SQ Sequence 284 AA;
 Query Match 15.5%; Score 217; DB 6; Length 284;
 Best Local Similarity 30.3%; Pred. No. 7.1e-13;
 Matches 70; Conservative 26; Mismatches 95; Indels 40; Gaps 9;
 QY 25 PMCQSVQNSFYKLSPPGLGQ-----LAAGTGHGTTILSRPVATPNSLLSGYPHV 77
 DB 33 PACLY-----MGRQPPPPPPQFTSLGSLGQSPDPISPYEVPLASDDPAGAHLLHHL 87
 QY 78 AGFGGLSSQGVYQVGSFSGKAGNEYPTTR-----NCWADTGDWRGSRAPC 126
 DB 88 PAQLGLAHPPP--GFPNGTEPGGLEENRVLPFPWMKSTKAHAW--KGQ-WAGGA--- 139
 QY 127 GNTDPDLSDTIHKKKHTRPTFTGHQIFALEKTFEQTKYLAGPERARLAYSIGMTESQVKV 186
 DB 140 -YTAEP-----EENKTRTAYTRAQILLEKEFLFNKYSRPRVELAVMLNLTHERHIKI 193
 QY 187 WFNQRRTKWRKKSALPSSSTPRAPGGA---SGDRAASENEDDEYNKPLDP 234
 DB 194 WFNQRRMKWKEEDKKRSSGTSPGGGGEEPEQDCAVTSGBELLAVPPLPP 244
 RESULT 19
 ID ADF89562 standard; protein; 284 AA.
 XX
 XX ADF89562;
 AC ADF89562;
 XX
 DT 26-FEB-2004 (first entry)

XX Mouse pdx-1 protein.
DE
XX insulin production cell; pancreatic beta-cell transcription factor;
KW pdx-1; neurogenin-3; wip common-bile-duct administration;
KW type I diabetes; mouse; murine.
XX
XX Mus musculus.
OS
XX JP2003189875-A.
FN
XX 08-JUL-2003.
PD
XX 28-DEC-2001; 2001JP-00399251.
PF
XX 28-DEC-2001; 2001JP-00399251.
PR
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA
XX WPI; 2003-819453/77.
DR N-PSDB; ADF89561.
DR
XX Inducing insulin production cell formation useful for treating diabetes
PT involves introducing gene encoding pancreatic beta cell-related
XX transcription factor to pancreas by wip common-bile-duct administration.
PT
XX Example; SEQ ID NO 2; 19pp; Japanese.
PS
XX The invention comprises a method for inducing an insulin production cell
CC formation, the method involves introducing a gene encoding a pancreatic
CC beta-cell-related transcription factor (e.g. pdx-1 or neurogenin-3) to the
CC pancreas by wip common-bile-duct administration. The method of the
CC invention is useful for inducing the formation of an insulin production
CC cell, and thus for treating diabetes - preferably type I diabetes. The
CC present amino acid sequence represents the mouse pdx-1 protein.
XX
XX Sequence 284 AA;
SQ
Query Match 15.5%; Score 217; DB 7; Length 284;
Best Local Similarity 30.3%; Pred. No. 7.1e-13;
Matches 70; Conservative 26; Mismatches 95; Indels 40; Gaps 9;
QY 25 PNCQSVQNSFYKLSPPGLGQ-----LAAGTGHGTDILSRVATPNSLLSGYPHV 77
Db 33 PACLY-----MGRQPPPPPPQFTSSLSGSGSPDISPYEVPLASDDPAGAHLLHL 87
QY 78 AGFGGLSQGVYQVQVGSFSGAGNEYPTRTR-----NCWADTGDWRGSRAPC 126
Db 88 PAQLGLAHPPP--GFFPNGTEPGGLEENRVLPPFPWPKSKAHAW--KGQ-WAGGA--- 139
QY 127 GNTDPPLSDTHKKHTRPTTGHQIFALEKTFQTKYLAGPERARLAYSLGTMESQVKV 186
Db 140 -YTAEP-----EENKRTTAVTRAQLLEKEFLPKYISRRVVELAVMLNTERHIKI 193
QY 187 WQNRRTWRKKSALPSSSTPRAPGGA---SGDPAASENEDDEYKPLDP 234
Db 194 WQNRRTWRKKSALPSSSTPRAPGGA---SGDPAASENEDDEYKPLDP 244

KW paired-box transcription factor 4; Nkx2.2; NK-2 class homeobox protein;
KW Nkx6.1; Isl-1; Islet factor 1; Pdx-1;
KW pancreatic and duodenal homeobox protein; BCT; betacellulin.
XX
XX Mus musculus.
OS
XX US2004132679-A1.
FN
XX 08-JUL-2004.
PD
XX 03-SEP-2003; 2003US-00654102.
PF
XX 03-SEP-2002; 2002US-0407743P.
PR
XX (BAYU) BAYLOR COLLEGE MEDICINE.
PA
XX Chan L, Kojima H;
PI
XX WPI; 2004-517032/49.
DR
XX Use of an islet cell differentiation transcription factor polypeptide or
PT its homologue or analog for treating a mammal for insulin-dependent
PT diabetes, increasing an insulin level in a somatic cell, or generating an
PT insulin-producing cell.
XX
PS Claim 109; SEQ ID NO 65; 190pp; English.
XX
XX The invention relates to the use of an islet cell differentiation
CC transcription factor polypeptide or its homologue or analogue for
CC treating a mammal for insulin-dependent diabetes, increasing an insulin
CC level in a somatic cell, or generating an insulin-producing cell. Also
CC included are a method of treating a mammal for insulin-dependent
CC diabetes, a method of increasing an insulin level in a somatic cell, a
CC method of generating an insulin-producing cell, a therapeutic composition
CC comprising an isolated islet cell differentiation transcription factor
CC polypeptide (and/or an isolated nucleic acid expressing the polypeptide),
CC an insulin-producing cell comprising a vector (where the vector comprises
CC a nucleic acid sequence encoding an islet cell differentiation
CC transcription factor), an insulin-producing cell (generated by a method
CC comprising obtaining a somatic cell and transfecting the cell with a
CC vector comprising a nucleic acid sequence encoding an islet cell
CC differentiation transcription factor, where in the transfecting step the
CC cell produces insulin), a method of generating at least one pancreatic
CC islet, and a composition (comprising: NeuroD or ngn3 polypeptide or a
CC polynucleotide expressing a NeuroD or ngn3 polypeptide and betacellulin
CC polypeptide or a polynucleotide expressing a betacellulin polypeptide).
CC The islet cell differentiation transcription factor polypeptide is
CC NeuroD, ngn3 (neurogenin 3), Pax6 (paired-box transcription factor 6),
CC Pax4 (paired-box transcription factor 4), Nkx2.2 (NK-2 class homeobox
CC protein), Nkx6.1, Isl-1 (Islet factor 1), Pdx-1 (pancreatic and duodenal
CC homeobox protein), BCT (betacellulin) or their combinations. The islet
CC cell differentiation transcription factor polypeptide or its homologue or
CC analogue is useful for treating a mammal for insulin-dependent diabetes,
CC increasing an insulin level in a somatic cell, or generating an insulin-
CC producing cell. The present sequence is an islet cell differentiation
CC transcription factor polypeptide as detailed above.
XX
XX Sequence 284 AA;
SQ
Query Match 15.5%; Score 217; DB 8; Length 284;
Best Local Similarity 30.3%; Pred. No. 7.1e-13;
Matches 70; Conservative 26; Mismatches 95; Indels 40; Gaps 9;
QY 25 PNCQSVQNSFYKLSPPGLGQ-----LAAGTGHGTDILSRVATPNSLLSGYPHV 77
Db 33 PACLY-----MGRQPPPPPPQFTSSLSGSGSPDISPYEVPLASDDPAGAHLLHL 87
QY 78 AGFGGLSQGVYQVQVGSFSGAGNEYPTRTR-----NCWADTGDWRGSRAPC 126
Db 88 PAQLGLAHPPP--GFFPNGTEPGGLEENRVLPPFPWPKSKAHAW--KGQ-WAGGA--- 139
QY 127 GNTDPPLSDTHKKHTRPTTGHQIFALEKTFQTKYLAGPERARLAYSLGTMESQVKV 186
XX

Db 140 -YTAEP-----BENKRTTAYTRAQLELEKEFLFNKYSRPRRVELAVMLNLTERHIKI 193
 QY 187 WFNQRRTKWRKKSALPSSSTRAPGGA---SGDRAASENEDDEYNKPLDP 234
 Db 194 WFNQRRMKWKEEDKRRSGTSPGGGGEPEPDQCAVTSGBELLAVPPLPP 244

RESULT 21

ADQ09818
 ID ADQ09818 standard; protein; 284 AA.
 XX
 AC ADQ09818;
 DT 23-SEP-2004 (first entry)
 XX Mouse pancreatic and duodenal homeobox protein, Pdx-1, #1.
 DE
 DE Mouse; islet cell differentiation transcription factor;
 KW insulin-dependent diabetes; insulin; somatic cell;
 KW insulin-producing cell; islet cell; NeuroD; ngn3; neurogenin 3; Pax6;
 KW paired-box transcription factor 6; Pax4;
 KW paired-box transcription factor 4; Nkx2.2; NK-2 class homeobox protein;
 KW Nkx6.1; Isl-1; Islet factor 1; Pdx-1;
 KW pancreatic and duodenal homeobox protein; BCT; betacellulin.
 XX
 OS Mus musculus.
 XX
 XX US2004132679-A1.
 XX
 PD 08-JUL-2004.
 XX
 PF 03-SEP-2003; 2003US-00654102.
 XX
 PF 03-SEP-2002; 2002US-0407743P.
 XX
 PR (BAYU) BAYLOR COLLEGE MEDICINE.
 XX
 PA Chan L, Kojima H;
 PI WPI: 2004-517032/49.
 DR N-PSDB; ADQ09876, ADQ09853.
 XX

Use of an islet cell differentiation transcription factor polypeptide or its homologue or analog for treating a mammal for insulin-dependent diabetes, increasing an insulin level in a somatic cell, or generating an insulin-producing cell.
 PS Claim 109; SEQ ID NO 56; 190pp; English.

The invention relates to the use of an islet cell differentiation transcription factor polypeptide or its homologue or analogue for treating a mammal for insulin-dependent diabetes, increasing an insulin level in a somatic cell, or generating an insulin-producing cell. Also included are a method of treating a mammal for insulin-dependent diabetes, a method of increasing an insulin level in a somatic cell, a method of generating an insulin-producing cell, a therapeutic composition comprising an isolated islet cell differentiation transcription factor polypeptide (and/or an isolated nucleic acid expressing the polypeptide), an insulin-producing cell comprising a vector (where the vector comprises a nucleic acid sequence encoding an islet cell differentiation transcription factor), an insulin-producing cell (generated by a method comprising obtaining a somatic cell and transfecting the cell with a vector comprising a nucleic acid sequence encoding an islet cell differentiation transcription factor, where in the transfecting step the cell produces insulin), a method of generating at least one pancreatic islet, and a composition (comprising: NeuroD or ngn3 polypeptide or a polynucleotide expressing a NeuroD or ngn3 polypeptide and betacellulin polypeptide or a polynucleotide expressing a betacellulin polypeptide). The islet cell differentiation transcription factor polypeptide is NeuroD, ngn3 (neurogenin 3), Pax6 (paired-box transcription factor 6), Pax4 (paired-box transcription factor 4), Nkx2.2 (NK-2 class homeobox protein), Nkx6.1, Isl-1 (Islet factor 1), Pdx-1 (pancreatic and duodenal homeobox protein), BCT (betacellulin) or their combinations. The islet

cell differentiation transcription factor polypeptide or its homologue or analogue is useful for treating a mammal for insulin-dependent diabetes, increasing an insulin level in a somatic cell, or generating an insulin-producing cell. The present sequence is an islet cell differentiation transcription factor polypeptide as detailed above.
 XX
 SQ Sequence 284 AA;

Query Match 15.5%; Score 217; DB 8; Length 284;
 Best Local Similarity 30.3%; Pred. No. 7.1e-13;
 Matches 70; Conservative 26; Mismatches 95; Indels 40; Gaps 9;
 QY 25 PMSQVSVQNSFYKLSPPGLGPQ-----LAAGTGHGTTDILSRPVATPNSLLSGYPHV 77
 Db 33 PACLY-----MGRQPPPPPPPPFTSSLSGLSQGPPDISPYEVPLASDDPAGHLHHHL 87
 QY 78 AGFGGLSSQGVYGPQVGSFSGKAGNEYPTTR-----NCWADTGDMWGSARPC 126
 Db 88 PAQLGLAHPPP--GPFNGTEPGGLEPFRVQLPFPWMKSTKAHAW--KGQ-WAGGA--- 139
 QY 127 GNTDPPLSDTIHKKHTRPTTGHQIIFALEKTFQTKYLACPERARLAYSLGTMESQVKV 186
 Db 140 -YTAEP-----BENKRTTAYTRAQLELEKEFLFNKYSRPRRVELAVMLNLTERHIKI 193
 QY 187 WFNQRRTKWRKKSALPSSSTRAPGGA---SGDRAASENEDDEYNKPLDP 234
 Db 194 WFNQRRMKWKEEDKRRSGTSPGGGGEPEPDQCAVTSGBELLAVPPLPP 244

RESULT 22

ADQ09823
 ID ADQ09823 standard; protein; 284 AA.
 XX
 AC ADQ09823;
 XX
 DT 23-SEP-2004 (first entry)
 XX Mouse pancreatic and duodenal homeobox protein, Pdx-1, #2.
 DE
 DE Mouse; islet cell differentiation transcription factor;
 KW insulin-dependent diabetes; insulin; somatic cell;
 KW insulin-producing cell; islet cell; NeuroD; ngn3; neurogenin 3; Pax6;
 KW paired-box transcription factor 6; Pax4;
 KW paired-box transcription factor 4; Nkx2.2; NK-2 class homeobox protein;
 KW Nkx6.1; Isl-1; Islet factor 1; Pdx-1;
 KW pancreatic and duodenal homeobox protein; BCT; betacellulin.
 XX
 OS Mus musculus.
 XX
 XX US2004132679-A1.
 XX
 PD 08-JUL-2004.
 XX
 PF 03-SEP-2003; 2003US-00654102.
 XX
 PF 03-SEP-2002; 2002US-0407743P.
 XX
 PR (BAYU) BAYLOR COLLEGE MEDICINE.
 XX
 PA Chan L, Kojima H;
 PI WPI: 2004-517032/49.
 DR
 DR Use of an islet cell differentiation transcription factor polypeptide or its homologue or analog for treating a mammal for insulin-dependent diabetes, increasing an insulin level in a somatic cell, or generating an insulin-producing cell.
 XX
 PS Claim 109; SEQ ID NO 61; 190pp; English.
 XX
 XX The invention relates to the use of an islet cell differentiation transcription factor polypeptide or its homologue or analogue for treating a mammal for insulin-dependent diabetes, increasing an insulin level in a somatic cell, or generating an insulin-producing cell.

QY 12 NNTQLAQSEMKAPMCQSVQNSFYKLS-----PPGLGQPLAAGTPHGTIDILSRVATPN 67
 Db 288 NHQALQHQOQSWPPRHYS--GSWYPTSLSEIPISAPNIASTVAYASGPSLAHSLSPN 345
 QY 68 SSSLGYPHVAGFGLSGQVYVYGVQVGSFKAGNEYPTRTNRCWADT----- 115
 Db 346 -----DIESLASIGHQ-----RNCVATEDIHLKELDGH 375
 QY 116 -----GQWRGSAAPCGNTDPPLSDTIHKK--HTRPTFTGHQIFALEKTEPQTKY 164
 Db 376 QSDGTSGEGENSGASNIGNTEDDQARLILKRLQRNRTSFTNDQIDSLKEPERTHY 435
 QY 145 LAGPRARLAVSLGMTESQVQVWFONRTKWKKSALPSSSTPRAPGASGDRAAS 221
 Db 436 PDVFARERLAGKIGLPEARIQVWFNSRAKWRREKLRNQRRTPNSTGASATSSSTS 492

RESULT 26
 ADA83730
 ID ADA83730 standard; protein; 328 AA.
 XX
 AC ADA83730;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Human DLX2 protein.
 XX
 KW human; marker; expressed sequence tag; EST; arabidopsis; tumour;
 KW stress-induced phenotype; hyperosmotic stress; colon cancer; immunogen;
 KW vaccine.
 XX
 OS Homo sapiens.
 XX
 PN WO2002103028-A2.
 XX
 PD 27-DEC-2002.
 XX
 PF 30-MAY-2002; 2002WO-IB004189.
 XX
 PR 30-MAY-2001; 2001US-02939999.
 PR 22-OCT-2001; 2001US-0330457P.
 PR 19-FEB-2002; 2002US-0357144P.
 XX
 PA (BIOM-) BIOMEDICAL CENT.
 XX
 PI Baranova AV, Yankovsky NK, Kozlov AP, Lobashev AV, Krukovskaya LL;
 XX
 XX WPI; 2003-175241/17.
 DR N-PSDB; ADA83729.
 XX
 XX Determining if a nucleic acid is a marker for a phenotype/cell type of
 PT interest, by global comparison of expressed sequence tags known to be
 PT expressed in the phenotype/cell type with all ESTs expressed in normal
 PT tissue.
 XX
 XX Claim 29; Page 71-72; 516pp; English.
 PS
 XX The invention relates to a novel method for determining if a nucleic acid
 CC is a marker for a predetermined phenotype/cell type of interest from a
 CC biological species. The method comprises performing a global comparison
 CC of a group of expressed sequence tags (ESTs) known to be expressed in the
 CC phenotype/cell type of interest with all ESTs expressed in normal tissue
 CC in order to identify ESTs that are preferentially expressed in the
 CC phenotype/cell of interest. A method of the invention is useful for
 CC determining whether a nucleic acid is a marker for a predetermined
 CC phenotype or cell type of interest from a biological species, preferably
 CC Arabidopsis or human. The cell type of interest is an abnormal cell such
 CC as a tumour cell, and the predetermined phenotype is a stress-induced
 CC phenotype such as hyperosmotic stress or high salt conditions. A method
 CC of the invention is also useful for determining the progression of colon
 CC cancer in a human, for detecting a tumour cell, and for regulating or
 CC preventing the growth of a tumour cell. An antibody of the invention is

CC useful for detecting the absence or presence of peptides encoded by
 CC tumour-associated markers. A polypeptide of the invention is useful as an
 CC immunogen for vaccinating an animal. The present sequence represents a
 CC tumour-associated antigen of the invention.
 XX
 SQ Sequence 328 AA;
 Query Match 15.2%; Score 212.5; DB 6; Length 328;
 Best Local Similarity 28.9%; Pred. No. 2.5e-12;
 Matches 68; Conservative 42; Mismatches 86; Indels 39; Gaps 10;
 QY 19 FSEMKAPM--COYSQVNSFYKLSPPGLGQPLAAGTPHGTIDILSRVATP-----NS 68
 Db 5 FDSLAVDMHSTQIAASSTYHQHQPSPGCGAGPGNSSSSSLHKKPQSPPTLPVSTATDS 64
 QY 69 SLLSGYPHVAGFGLSGQVYVYGVQVGSF---SKAGNEYPTRTNRCWADTG-----Q 117
 Db 65 SYVTNQHPAGGGG--GGSPYA-HMGSYQVQASGLNNVPYSKSSY-DLGYYTAAYTVA 120
 QY 118 DWGSAAPCGNTDP-----LSDTIHKKHTPTFTGHQIFALEKTEPQTKYLAP 168
 Db 121 PYGTSSSPANNEPEKEDLEPEIRIVNGKPKVKRPTTIYSSFFQLAALQRRFQKQYLA 180
 QY 169 ERARLAVSLGMTESQVQVWFONRTKWK--KSALEPSSSTPRAPGASGDRAAS 221
 Db 181 ERAELASLGLTQVQVWFONRRSKFKQMKWMSGEIPSE---QHPGASAPPCAS 232

RESULT 27
 AAW86020
 ID AAW86020 standard; protein; 330 AA.
 XX
 AC AAW86020;
 XX
 DT 22-FEB-1999 (first entry)
 XX
 DE Human homeobox gene HOX11 protein sequence.
 XX
 KW HOX11 gene; homeobox; transgenic; mouse; human; gene inactivation;
 KW asplenic; screening; developmental; pathological; thymus; T cell;
 KW organogenesis.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FH Peptide 19..25
 FT /note= "help motif"
 FT Domain 200..260
 FT /note= "homeobox domain"
 XX
 XX US950002-A.
 XX
 XX 15-DEC-1998.
 XX
 PF 13-SEP-1996; 96US-00712948.
 XX
 PR 20-APR-1994; 94US-00231728.
 XX
 XX (UNIW) UNIV WASHINGTON.
 XX
 XX Korameyer SJ;
 XX
 XX WPI; 1999-069841/06.
 XX N-PSDB; AAW86020.
 XX
 XX Transgenic mouse with inactivated Hox11 gene - and transgenic mouse
 PT having human Hox11 gene, used for screening compounds.
 XX
 XX Disclosure; Col 19-20; 26pp; English.
 XX
 XX The invention provides a transgenic mouse homozygous for Hox11 gene
 CC inactivation. The Hox11 gene is inactivated by the insertion of a
 CC heterologous nucleic acid sequence into the Hox11 gene via homologous

CC recombination, where the inactivation of the Hox11 gene prevents
 CC expression of the Hox11 gene, and whereby the mouse is asplenic. The
 CC mouse is used for screening for a compound, where the compound affects a
 CC developmental and/or pathological condition due to expression of the
 CC human Hox11 gene in the thymus or T cells that do not normally express
 CC the human Hox11 gene. Animals deficient in orphan homeobox gene function
 CC serve as models for organogenesis. The present sequence represents the
 CC human homeobox Hox11 protein sequence
 XX
 SQ Sequence 330 AA;

Query Match 15.2%; Score 212.5; DB 2; Length 330;
 Best Local Similarity 30.2%; Pred. No. 2.5e-12;
 Matches 74; Conservative 29; Mismatches 73; Indels 69; Gaps 11;
 QY 36 YKLSPPGLPQLAAGTGHGIDILSRPVATPNSSL---LSGYPHVAGFGG-----LSS 85
 DB 74 YTGTFGG-GP-----GPGAGGGGACSMGLTGSYNVMALAGPGGGGGSSGGAGALSA 128
 QY 86 QGVY----YGPQVGSFSK-----AGNEYPTTRNCWADTQD 118
 DB 129 AGVIRVPAHRPLAGVAHQPLATGLTPVSPAMPVGNVNLTLTFP-----WVESNR 182
 QY 119 W---RGSARPCONTDDSDTHKKKTRPTTGHQIFALEKTPFQTKYLACPERARLAY 175
 DB 183 YTKDRFTGHYPQNTFP-----KKKPRTSFTRLQICELEKRFHQKYLASAERAAAK 236
 QY 176 SLGMTESQVKVWFQNRRTKWKKSALPSSSTPRAPGASGDRAASENEDDEYNK----- 230
 DB 237 ALKMTDAQVKVWFQNRRTKWKKSALPSSSTPRAPGASGDRAASENEDDEYNK----- 230
 QY 231 -PLDP 234
 DB 291 LPADP 295

RESULT 28
 AA083396
 ID AA083396 standard; protein; 284 AA.
 AC AA083396;
 XX
 DT 22-FEB-1999 (first entry)
 DE Mouse insulin promoter factor 1.
 KW Mouse; insulin promoter factor 1; transcriptional regulatory factor;
 KW Ipfl; drug screening; proliferation; differentiation; tissue maintenance.
 OS Mus sp.

Key Location/Qualifiers
 FT Misc-difference 16 /note= "encoded by GAC"
 FT Misc-difference 159 /note= "encoded by CAC"
 FT Misc-difference 160 /note= "encoded by CTG"
 FT Misc-difference 160 /note= "encoded by CTG"

US5849989-A.
 PD 15-DEC-1998.
 XX
 XX 07-OCT-1994; 94US-00320148.
 XX
 XX 07-OCT-1994; 94US-00320148.
 XX (ONTO-) ONTOGENY INC.

Edlund T;
 WPI; 1999-069836/06.
 DR N-FSDB; AAV72896.

XX Transgenic mouse including disrupted gene for insulin promoter factor 1 -
 PT lacking functional pancreas, used as model for cellular diseases and for
 PT drug screening.

PS Example 1; Col 51-52; 33pp; English.

XX The present invention describes a transgenic mouse having its genome
 CC altered to include a heterologous nucleic acid that, by homologous
 CC recombination with an endogenous ipfl (insulin promoter factor-1) gene,
 CC inhibits expression of the ipfl gene so that the mouse lacks an insulin-
 CC producing pancreas. Also described are transgenic mice in which
 CC expression of one allele of ipfl is inhibited by homologous recombination
 CC and which can be bred to produce homozygotes without a pancreas. The mice
 CC are useful as models for: (1) studying cellular diseases that involve
 CC mutated or misexpressed ipfl alleles i.e. for drug screening; and (2) for
 CC studying the role of ipfl in proliferation, differentiation and
 CC maintenance of tissue in vivo. The present sequence represents ipfl,
 CC which is used in an example from the present invention for cloning and
 CC expressing Ipfl
 XX Sequence 284 AA;

Query Match 15.1%; Score 211; DB 2; Length 284;
 Best Local Similarity 30.3%; Pred. No. 2.8e-12;
 Matches 70; Conservative 25; Mismatches 96; Indels 40; Gaps 9;
 QY 25 PMCOYSVQNSFYKLSPPCLGPQ-----LAAGTGHGIDILSRPVATPNSLSGYPHV 77
 DB 33 PACLY-----MGROPPPPPPPPFTSSLSGSLQSGSPDISPYEVPFLASDDPAGAHLLHH 87
 QY 78 AGFGGLSSQGVYGVQVGSFSGACNEYPTRTR-----NCWADTGDWGSARPC 126
 DB 88 PAQLGLAHPPP--GPFNGTEPGLEFPNVRVQLPFPWPKSTKAHAW--KGQ-WAGGA--- 139
 QY 127 GNTDPLSDTHKKKTRPTTGHQIFALEKTPFQTKYLACPERARLAYSLGMTESQVKV 186
 DB 140 -YTAEP-----EENKTRTAYTRAQSSSELEKEFLNKYISPRRVELAVMLNTERHIKI 193
 QY 187 WFNORRTKWKKSALPSSSTPRAPGGA---SGDRAASENEDDEYNKPLDP 234
 DB 194 WFNORRTKWKKEDEKRSSTGTPSGGGGEBEPQCAVTSGEELLAVPPLPP 244

RESULT 29
 AA099330
 ID AA099330 standard; protein; 284 AA.
 AC AA099330;
 XX
 DT 24-SEP-2002 (first entry)
 DE Mouse insulin promoter factor 1 (Ipfl) protein, version #2.
 KW Mouse; cytostatic; insulin promoter factor 1; Ipfl; diabetes mellitus;
 KW metabolic disorder; blood glucose; hyperglycaemia; insulin;
 KW transcription regulatory factor; Ipfl-responsive element; Ipfl-RE;
 KW P1 insulin promoter; immunogen; antisense; gene therapy;
 KW cell proliferation; cell differentiation; peptide hormone; cancer.

Mus sp.
 Key Location/Qualifiers
 FT Misc-difference 159.160 /note= "Encoded by CTGCTG"
 FT
 XX US2002082410-A1.
 PD 27-JUN-2002.

12-JAN-2001; 2001US-00759847.
 07-OCT-1994; 94US-00320148.

Db 194 WFNRRMKWKBEDEKRRSSGTFSGGGGEBEPQDCAVTSGBELLAVPPLPP 244

RESULT 31

ABB61602
ID ABB61602 standard; protein; 544 AA.

XX AC ABB61602;
XX

DT 26-MAR-2002 (first entry)
XX

DE Drosophila melanogaster polypeptide SEQ ID NO 11598.
XX

KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.

OS Drosophila melanogaster.
XX

FN WO200171042-A2.
XX

PD 27-SEP-2001.
XX

PF 23-MAR-2001; 2001WO-US009231.
XX

PR 23-MAR-2000; 2000US-0191637P.
XX

PR 11-JUL-2000; 2000US-00614150.
XX

PA (PEKE) PE CORP NY.
XX

PI Venter JC, Adams M, Li PWD, Myers EW;
XX

DR WPI; 2001-656860/75.
XX

DR N-PSDB; ABL05705.
XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.

PS Disclosure; SEQ ID NO 11598; 21pp + Sequence Listing; English.
XX

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 544 AA;

Query Match 15.1%; Score 210.5; DB 4; Length 544;

Best Local Similarity 29.0%; Pred. No. 8.2e-12;

Matches 54; Conservative 33; Mismatches 74; Indels 25; Gaps 4;

QY 40 PPGLG-----PQLAAGTPHGITDILSRVPATPNSLLSGYPHVAGFGLSSQ----- 86

Db 176 PPTAGGNGLNVAQYAAAHQHVAAAAAARNNAAAAAAGVAAAPVDDGVD 235

QY 87 -GVYVGPOVGSFKAGNYPTTRNC-----WADTQDWRGSRAPCGNTDPPLSD 135

Db 236 GGVGLAPPAGGLDSDSYHEENEDCSGNMDHSCVCSNGGKDDGNSVKSGTSD-MSG 294

QY 136 TIHKKKHTRPTTGHQIFALEKTFEOTKYLAGPERARLAYSLGWTESQVKWVFQNRRTKW 195

Db 295 LSKKKQKARTAPTTHQLOTLEKSFQKYLVSQERQELAHKLDLSDCQVKWYQNRRTKW 354

QY 196 RKKSAL 201

Db 355 KRQTAV 360

RESULT 32

ABB63899
ID ABB63899 standard; protein; 722 AA.

XX AC ABB63899;
XX

DT 26-MAR-2002 (first entry)
XX

DE Drosophila melanogaster polypeptide SEQ ID NO 18489.
XX

KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.

OS Drosophila melanogaster.
XX

FN WO200171042-A2.
XX

PD 27-SEP-2001.
XX

PF 23-MAR-2001; 2001WO-US009231.
XX

PR 23-MAR-2000; 2000US-0191637P.
XX

PR 11-JUL-2000; 2000US-00614150.
XX

PA (PEKE) PE CORP NY.
XX

PI Venter JC, Adams M, Li PWD, Myers EW;
XX

DR WPI; 2001-656860/75.
XX

DR N-PSDB; ABL08002.
XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.

PS Disclosure; SEQ ID NO 18489; 21pp + Sequence Listing; English.
XX

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 722 AA;

Query Match 15.0%; Score 209.5; DB 4; Length 722;

Best Local Similarity 32.7%; Pred. No. 1.6e-11;

Matches 70; Conservative 20; Mismatches 63; Indels 61; Gaps 9;

QY 37 KLS-----PPGLGQLAAGTPHGITDILSRVPATPNSLLSGYPHVAGFGLSSQGVYVG 91

Db 327 KLSVTALPPDI-----SPTGSSDSLMDKLMANN-----SSPGSNVN 365

QY 92 PQVGSFSKAGNEYPTTRNCWADTQDWRGSRAPCGNTDPPLSDTIHKKKHTRPTTGHQ 151

Db 366 AQMS-----NANSTLETED-----DSDSGSTDARRKKKARTTFTGRQ 404

QY 152 IFALEKTFEOTKYLAGPERARLAYSLGWTES-QVKWVFQNRRTKWRK-----SALEPS 204

Db 405 IFELEKTFENKYLISASERTEMAKLLVTTETQQVKIWFQNRRTKWKQDNVTNNEAEHK 464

QY 205 SSTPRAPCGSGDRAASENEDDEYNKFLDPDSD 238

Db 465 SS--NAKPGATGTATTTPS-----GEPTDKSSN 491

RESULT 33

AB061848
ID ABB61848 standard; protein; 263 AA.
XX
AC ABB61848;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 12336.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NV.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL05951.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 12336; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 263 AA;
Query Match 15.0%; Score 209; DB 4; Length 263;
Best Local Similarity 37.0%; Pred. No. 4e-12;
Matches 60; Conservative 18; Mismatches 48; Indels 36; Gaps 7;
QY 52 PHGTDILSRPVATPNSILSQ-----YPHVAGFGLSSQGVYGPQVGSFS-KAG 101
DB 65 PQHT-----TPTSSSSGGSSLTTHPHPLTG-----SHGGYLLPSSNESDEG 111
QY 102 NEYPTTRNCWADTGDW-RGSARPCGNTPDPLSDTIHKKKHTRFTTGHQIFALEKTFE 160
DB 112 EEIIEE-----DDGTGDSSSPHGD-----GNSRKKKTRTVFSRAQVQLESTFD 159
QY 161 QTKYLAGPERARLAYSLGWTESQVWFQNRRTKRWKSALE 202
DB 160 LKRYLSSSERAGLAASRLTETQVKIWFQNRNRNKKRQLAAE 201
RESULT 34
ID AAY84616
XX
AC AAY84616 standard; protein; 300 AA.
XX
DT 25-JUL-2000 (first entry)

XX
DE A chicken motor neuron restricted pattern 2 protein.
XX
KW Motor neuron restricted pattern 2 protein; MNR2 protein; motor neuron;
KW Pax6+ motor neuron progenitor; somatic motor neuron;
KW transcription factor; Isl2; Lim 3; homeobox protein 9; HB9;
KW chronic neurodegenerative disease; spinal muscular atrophy;
KW myotrophic lateral sclerosis; neuromuscular disease; sacral agenesis.
XX
OS Gallus sp.
XX
PN WO200018884-A1.
XX
PD 06-APR-2000.
XX
PF 29-SEP-1999; 99WO-US022517.
XX
PR 29-SEP-1998; 98US-00162524.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Jessell TM, Tanabe Y, William C;
XX
DR WPI; 2000-293133/25.
DR N-PSDB; AAZ99988.
XX
PT New isolated nucleic acid molecules encoding motor neuron restricted
PT pattern protein and a homeobox protein are used in treating chronic
PT neurodegenerative diseases and acute nervous system injuries.
XX
PS Claim 31; Fig 10; 209pp; English.
XX
CC The present sequence represents a motor neuron restricted pattern (MNR2)
CC protein from a chicken embryo. MNR2 is expressed selectively by Pax6+
CC motor neuron progenitors and persists in post-mitotic somatic motor
CC neurons. Expression of the MNR2 protein induces expression of
CC transcription factors Isl2, Lim 3 and homeobox protein 9 (HB9). MNR2 and
CC HB9 are used to induce differentiation of somatic motor neurons in a
CC mammal which is to treat an abnormality associated with a lack of one or
CC more normally functioning motor neurons, a chronic neurodegenerative
CC disease such as spinal muscular atrophy or myotrophic lateral sclerosis,
CC an acute nervous system injury localized to a specific central axon and
CC neuromuscular disease. HB9 is also used to treat an embryo afflicted with
CC sacral agenesis
XX
SQ Sequence 300 AA;
Query Match 14.9%; Score 208; DB 3; Length 300;
Best Local Similarity 27.0%; Pred. No. 6.1e-12;
Matches 75; Conservative 31; Mismatches 98; Indels 74; Gaps 10;
QY 39 SPPGLGQLAAGTGHGTDILSRPVATPNSILS-----GYPHVAGFGLSSQGVY 90
DB 26 SPPGLSPAGSPG-PAGRTDTPS-PPAPQAAATPLGPAFPVPEGLLHPG-PGLGLTLPALY 82
QY 91 GPQVGSFSKAGNEY-----PRTTRNCWADTG-----QDWRGSRPCGNTPD-- 131
DB 83 PPAVYPLPALGGQHAAPAYATAPQPPGGAHLLKAAVAGSPFLEQWIRAGMLVPRLSDFH 142
QY 132 --PLSDTIHKKKHTRFTTGHQIFALEKTFBQTKYLAGPERARLAYSLGWTESQVWFQ 189
DB 143 ATPQSALMGKSRPRRTAFTSQQLLENOFKLKYLSRPKPEVATSLMLTQVKIWFQ 202
QY 190 NRRTKWRKK-----SALEP-----SSSTPRAPGASG----- 216
DB 203 NRRMKKSRKAKEGMAVEPEKPRGLGKADESLLPSQPGQAGDSPEFVGSPGTGFLC 262
QY 217 -----DRAASENDEYNKPLDPDSDDEKIRLL 245
DB 263 RSAELGYDPDSSCGGEDEEEDDDGMDTAERKMGSVL 300

RESULT 35

XX The present sequence is that of new human protein regulating gene
 CC expression PRGE-3. It was deduced from incyte clone 996352 obtained from
 CC a kidney tumour cDNA library. PRGE-3 is characterised as a homeodomain
 CC protein. It is expressed in reproductive, cardiovascular and urologic
 CC tissues associated with cancer and trauma diseases, disorders or
 CC conditions. The invention provides PRGE polypeptides (see AAY58608-38)
 CC and polynucleotides (see AAZ57839-69), expression vectors, host cells,
 CC antibodies, agonists and antagonists. It also provides methods for
 CC diagnosing, treating or preventing disorders associated with expression
 CC of PRGE
 XX
 SQ Sequence 230 AA;
 Query Match 14.8%; Score 206; DB 3; Length 230;
 Best Local Similarity 25.3%; Pred. No. 6.6e-12;
 Matches 66; Conservative 34; Mismatches 71; Indels 90; Gaps 11;
 QY 19 FSEMKAPMCQY--SVQNSFYKLSPPGLPQLAAGTGHGTTDILSRPVATPNSLLSGYPH 76
 DB 21 FQNAEPTSCSPAPNSQSGY-----GAGAGAFASTVPLGLYV-----NSPLYQS-PF 66
 QY 77 VAGFG-----GLSS-----QGVYGPQVGSFSGAGNEYPT 106
 DB 67 ASGYGLGADAYGNLPCASYDQNIPLCLSDLAKGACDKTDEGALHGAEEANF-----RIYP- 121
 QY 107 RTRNCWADTGDWRGSRAPCGNTDPLSDTTHKKKHTRPTTGHQIFALEKTFQTKYLA 166
 DB 122 -----WMRSSGP-----DRKGRQTVTRYQTLEKEFHFNRYLI 156
 QY 167 GPERARLAYSIGMTESOVKVFQNRRTKWRKK-----SALEPSSSTPRAGGASGDRA 219
 DB 157 RRRIBIAHALCLTERQIKIWFQNRNMKWKKEHDEGPTAAAPGAVPSAAATAADKA 216
 QY 220 ASENEDEYKPLDPSDDEK 240
 DB 217 DEEDDEE-----EEDEEE 230
 RESULT 37
 ADQ17197
 ID ADQ17197 standard; protein; 291 AA.
 XX
 AC ADQ17197;
 XX
 DT 26-AUG-2004 (first entry)
 XX
 DE Human soft tissue sarcoma-upregulated protein - SEQ ID 13.
 XX
 KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
 XX
 OS Homo sapiens.
 XX
 PN WO2004048938-A2.
 XX
 PD 10-JUN-2004.
 XX
 PF 26-NOV-2003; 2003WO-US038193.
 XX
 PR 26-NOV-2002; 2002US-0429739P.
 XX
 PA (PROT-) PROTEIN DESIGN LABS INC.
 XX
 PI Aziz N, Ginsburg WM, Zlotnik A;
 XX
 DR WPI; 2004-441208/41.
 XX
 PT Early detection of soft tissue sarcoma comprises determining expression
 PT of a gene in a first soft tissue sample and a normal soft tissue sample
 PT and comparing the gene expression, also useful in treating soft tissue
 PT sarcoma.
 XX
 PS Example 2; SEQ ID NO 13; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma
 CC which comprises obtaining a first soft tissue sample from an individual
 CC and a normal soft tissue sample from the same or different individual,
 CC determining the expression of a gene in both samples and comparing the
 CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytosstatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC protein of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.
 XX
 SQ Sequence 291 AA;
 Query Match 14.8%; Score 206; DB 8; Length 291;
 Best Local Similarity 30.9%; Pred. No. 9.3e-12;
 Matches 75; Conservative 24; Mismatches 76; Indels 68; Gaps 11;
 QY 17 AQFSEMKAPM--COYSVQNSFYKLSPPGLPQLAAGTGHGTTDILSRPVATPNSLLSG 73
 DB 61 ASFAGPCGAPFEDAGSYSVNLS---LAPAGV-IRVPAHRP-----LPGAVPPPLPSALPA 110
 QY 74 YPHVAGFGGLSSQGVYGPQVGSFSGAGNEYPTTRNCWADTGDW----- 119
 DB 111 MPSV-----PTVSSLGRL--NFP-----WMSSRRFVKDRFTAAALTPF 148
 QY 120 ---RGSARPCGNTDPLSDTTHKKKHTRPTTGHQIFALEKTFQTKYLAGPERARLAYS 176
 DB 149 TVTRIRGHYPQNRTPP-----KXKPTPSRVSQICELEKRFHQKYLASAERAALAKS 202
 QY 177 LGMTESOVKVFQNRRTKWRKKSALEPSSSTPRAGGASGDRAASENEDDYKPLD--- 233
 DB 203 LKMTDAQKTVFQNRRTKWRRTAE-----READGQASRLMLQLQHDFAFKSLNDSI 256
 QY 234 -PD 235
 DB 257 QPD 259
 RESULT 38
 AAE10922
 ID AAE10922 standard; protein; 217 AA.
 XX
 AC AAE10922;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Human HOXB7 variant protein.
 XX
 KW Human; cytostatic; cancer-related antigen; homeobox protein; HOXA7;
 KW HOXB7; ADP ribosylation factor 1; Arf-1; ATP dependent iron transporter;
 KW ABC-7; neoplastic process; ovarian cancer; benign serous cystadenoma;
 KW vaccine; variant.
 XX
 OS Homo sapiens.
 XX
 PN WO200168853-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 14-MAR-2001; 2001WO-US007896.
 XX
 PR 14-MAR-2000; 2000US-0189226P.
 XX
 PR 28-DEC-2000; 2000US-0258452P.
 XX
 PA (UYUJ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
 XX
 PI Roden R, Naora H;
 XX
 DR WPI; 2001-596909/67.

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2005, 16:58:19 ; Search time 11.3345 Seconds
(without alignments)
375.348 Million cell updates/sec

Title: US-09-998-861-14
Perfect score: 11
Sequence: 1 XXXXXXXXXXXX 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	0	0.0	1	2	ADD95004 Platelet
2	0	0.0	1	2	ADD94992 Platelet
3	0	0.0	1	2	ADD94993 Platelet
4	0	0.0	1	2	ADD95002 Platelet
5	0	0.0	1	2	ADD94997 Platelet
6	0	0.0	1	2	ADD94999 Platelet
7	0	0.0	1	2	ADD95003 Platelet
8	0	0.0	1	2	ADD94995 Platelet
9	0	0.0	1	2	ADD94998 Platelet
10	0	0.0	1	2	ADD95001 Platelet
11	0	0.0	1	2	ADD94990 Platelet
12	0	0.0	1	2	ADD94991 Platelet
13	0	0.0	1	2	ADD94994 Platelet
14	0	0.0	1	2	ADD94996 Platelet
15	0	0.0	1	2	ADD95000 Platelet
16	0	0.0	1	2	AAV46652 Immunogen
17	0	0.0	1	2	AAV46652 Immunogen
18	0	0.0	1	4	AAV46652 Human pep
19	0	0.0	1	4	AAV46652 Human pep
20	0	0.0	1	4	AAV46652 Human pep
21	0	0.0	1	4	AAV46652 Human pep
22	0	0.0	1	4	AAV46652 Human pep
23	0	0.0	1	4	AAV46652 Human pep
24	0	0.0	1	4	AAV46652 Human pep
25	0	0.0	1	4	AAV46652 Human pep

26	0	0.0	1	4	AAV46652 Human pep
27	0	0.0	1	4	AAV46652 Human pep
28	0	0.0	1	4	AAV46652 Human pep
29	0	0.0	1	4	AAV46652 Human pep
30	0	0.0	1	4	AAV46652 Human pep
31	0	0.0	1	4	AAV46652 Human pep
32	0	0.0	1	4	AAV46652 Human pep
33	0	0.0	1	4	AAV46652 Human pep
34	0	0.0	1	4	AAV46652 Human pep
35	0	0.0	1	4	AAV46652 Human pep
36	0	0.0	1	4	AAV46652 Human pep
37	0	0.0	1	4	AAV46652 Human pep
38	0	0.0	1	4	AAV46652 Human pep
39	0	0.0	1	4	AAV46652 Human pep
40	0	0.0	1	4	AAV46652 Human pep
41	0	0.0	1	4	AAV46652 Human pep
42	0	0.0	1	4	AAV46652 Human pep
43	0	0.0	1	4	AAV46652 Human pep
44	0	0.0	1	4	AAV46652 Human pep
45	0	0.0	1	4	AAV46652 Human pep

ALIGNMENTS

RESULT 1
ADD95004
ID ADD95004 standard; peptide; 1 AA.
XX
AC ADD95004;
XX
DT 29-JAN-2004 (first entry)
XX
DE Platelet aggregation inhibitor peptide #146.
XX
KW platelet aggregation inhibitor; guanidino group; amidino group.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /label= OTHER
FT /note= "OTHER= H2NC(=NH)NH(CH2)CO. Beta-Ala (not defined)"
XX
FN WO9501371-A1.
XX
PD 12-JAN-1995.
XX
PF 22-JUN-1994; 94WO-JP000999.
XX
PR 30-JUN-1993; 93JP-00186755.
XX
PA (YAWA) NIPPON STEEL CORP.
XX
PI Sato Y, Hayashi Y, Katada J, Takiguchi Y;
XX WPI; 1995-060950/08.
XX
PT New RGD peptide(s) useful as anti-platelet aggregation agents - contain
XX guanidino or amidino gp. at N-terminal to increase stability.
XX
PS Disclosure; Page 11; 34pp; Japanese.
XX
CC The invention describes peptides of amino acid sequence (I) and their
XX salts. (I) are useful as platelet aggregation inhibitors and are easily
XX absorbed by the body. Due to the presence of the N-terminal guanidino or
XX amidino group, (I) are provided with excellent stability so that their
XX activity can be exhibited for an effective time after administration.
XX Thereafter they are readily metabolised and expelled. This is the amino
XX acid sequence of a platelet aggregation inhibitor peptide.
XX
SQ Sequence 1 AA;

Query Match 0.0%; Score 0; DB 2; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
 Db 1 R 1

RESULT 2
 ADD94992
 ID ADD94992 standard; peptide; 1 AA.

XX AC ADD94992;

DT 29-JAN-2004 (first entry)

DE Platelet aggregation inhibitor peptide #134.

KW platelet aggregation inhibitor; guanidino group; amidino group.

OS Unidentified.

XX FH Key Location/Qualifiers
 FT Modified-site 1
 FT /label= OTHER
 FT /note= "OTHER= H2NC(=NH)NH(CH2)CO. Beta-Ala (not
 defined)"

XX PN WO9501371-A1.

XX PD 12-JAN-1995.

PF 22-JUN-1994; 94WO-JP000999.

PR 30-JUN-1993; 93JP-00186755.

XX PA (YAWA) NIPPON STEEL CORP.

XX PI Sato Y, Hayashi Y, Katada J, Takiguchi Y;

XX DR WPI; 1995-060950/08.

XX PT New RGD peptide(s) useful as anti:platelet aggregation agents - contain
 guanidino or amidino gp. at N-terminal to increase stability.

XX PS Disclosure; Page 10; 34pp; Japanese.

XX CC The invention describes peptides of amino acid sequence (I) and their
 salts. (I) are useful as platelet aggregation inhibitors and are easily
 absorbed by the body. Due to the presence of the N-terminal guanidino or
 amidino group, (I) are provided with excellent stability so that their
 activity can be exhibited for an effective time after administration.
 CC Thereafter they are readily metabolised and expelled. This is the amino
 acid sequence of a platelet aggregation inhibitor peptide.

XX SQ Sequence 1 AA;

Query Match 0.0%; Score 0; DB 2; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
 Db 1 R 1

RESULT 3
 ADD94993
 ID ADD94993 standard; peptide; 1 AA.

XX AC ADD94993;

XX DT 29-JAN-2004 (first entry)
 XX DE Platelet aggregation inhibitor peptide #135.
 XX KW platelet aggregation inhibitor; guanidino group; amidino group.
 XX OS Unidentified.

XX FH Key Location/Qualifiers
 FT Modified-site 1
 FT /label= OTHER
 FT /note= "OTHER= H2NC(=NH)NH(CH2)CO. Beta-Ala (not
 defined)"

XX PN WO9501371-A1.

XX PD 12-JAN-1995.

PF 22-JUN-1994; 94WO-JP000999.

PR 30-JUN-1993; 93JP-00186755.

XX PA (YAWA) NIPPON STEEL CORP.

XX PI Sato Y, Hayashi Y, Katada J, Takiguchi Y;

XX DR WPI; 1995-060950/08.

XX PT New RGD peptide(s) useful as anti:platelet aggregation agents - contain
 guanidino or amidino gp. at N-terminal to increase stability.

XX PS Disclosure; Page 10; 34pp; Japanese.

XX CC The invention describes peptides of amino acid sequence (I) and their
 salts. (I) are useful as platelet aggregation inhibitors and are easily
 absorbed by the body. Due to the presence of the N-terminal guanidino or
 amidino group, (I) are provided with excellent stability so that their
 activity can be exhibited for an effective time after administration.
 CC Thereafter they are readily metabolised and expelled. This is the amino
 acid sequence of a platelet aggregation inhibitor peptide.

XX SQ Sequence 1 AA;

Query Match 0.0%; Score 0; DB 2; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
 Db 1 R 1

RESULT 4
 ADD95002
 ID ADD95002 standard; peptide; 1 AA.

XX AC ADD95002;

XX DT 29-JAN-2004 (first entry)

XX DE Platelet aggregation inhibitor peptide #144.

XX KW platelet aggregation inhibitor; guanidino group; amidino group.

XX OS Unidentified.

XX FH Key Location/Qualifiers
 FT Modified-site 1
 FT /label= OTHER
 FT /note= "OTHER= H2NC(=NH)NH(CH2)CO. Beta-Ala (not
 defined)"

Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 X 1
Db 1 R 1

RESULT 7
ADD95003
ID ADD95003 standard; peptide; 1 AA.
XX AC ADD95003;
XX 29-JAN-2004 (first entry)
XX Platelet aggregation inhibitor peptide #145.
XX platelet aggregation inhibitor; guanidino group; amidino group.
XX Unidentified.
XX Key Location/Qualifiers
FT Modified-site 1
FT /label= OTHER
FT /note= "OTHER= H2NC(=NH)NH(CH2)CO. Beta-Ala (not defined)"
XX WO9501371-A1.
XX 12-JAN-1995.
XX 22-JUN-1994; 94WO-JP0000999.
XX 30-JUN-1993; 93JP-00186755.
XX (YAWA) NIPPON STEEL CORP.
XX Sato Y, Hayashi Y, Katada J, Takiguchi Y;
XX WPI; 1995-060950/08.
XX New RGD peptide(s) useful as anti:platelet aggregation agents - contain guanidino or amidino gp. at N-terminal to increase stability.
XX Disclosure; Page 11; 34pp; Japanese.
XX The invention describes peptides of amino acid sequence (I) and their salts. (I) are useful as platelet aggregation inhibitors and are easily absorbed by the body. Due to the presence of the N-terminal guanidino or amidino group, (I) are provided with excellent stability so that their activity can be exhibited for an effective time after administration. Thereafter they are readily metabolised and expelled. This is the amino acid sequence of a platelet aggregation inhibitor peptide.
XX Sequence 1 AA;
Query Match 0.0%; Score 0; DB 2; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 X 1
Db 1 R 1

RESULT 8
ADD94995
ID ADD94995 standard; peptide; 1 AA.
XX AC ADD94995;
XX 29-JAN-2004 (first entry)

XX Platelet aggregation inhibitor peptide #137.
XX platelet aggregation inhibitor; guanidino group; amidino group.
XX Unidentified.
XX Key Location/Qualifiers
FT Modified-site 1
FT /label= OTHER
FT /note= "OTHER= H2NC(=NH)NH(CH2)CO. Beta-Ala (not defined)"
XX WO9501371-A1.
XX 12-JAN-1995.
XX 22-JUN-1994; 94WO-JP0000999.
XX 30-JUN-1993; 93JP-00186755.
XX (YAWA) NIPPON STEEL CORP.
XX Sato Y, Hayashi Y, Katada J, Takiguchi Y;
XX WPI; 1995-060950/08.
XX New RGD peptide(s) useful as anti:platelet aggregation agents - contain guanidino or amidino gp. at N-terminal to increase stability.
XX Disclosure; Page 10; 34pp; Japanese.
XX The invention describes peptides of amino acid sequence (I) and their salts. (I) are useful as platelet aggregation inhibitors and are easily absorbed by the body. Due to the presence of the N-terminal guanidino or amidino group, (I) are provided with excellent stability so that their activity can be exhibited for an effective time after administration. Thereafter they are readily metabolised and expelled. This is the amino acid sequence of a platelet aggregation inhibitor peptide.
XX Sequence 1 AA;
Query Match 0.0%; Score 0; DB 2; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 X 1
Db 1 R 1

RESULT 9
ADD94998
ID ADD94998 standard; peptide; 1 AA.
XX AC ADD94998;
XX 29-JAN-2004 (first entry)
XX Platelet aggregation inhibitor peptide #140.
XX platelet aggregation inhibitor; guanidino group; amidino group.
XX Unidentified.
XX Key Location/Qualifiers
FT Modified-site 1
FT /label= OTHER
FT /note= "OTHER= H2NC(=NH)NH(CH2)CO. Beta-Ala (not defined)"
XX WO9501371-A1.
XX


```

XX KW platelet aggregation inhibitor; guanidino group; amidino group.
XX OS Unidentified.
XX FH Key Location/Qualifiers
XX FT Modified-site 1
XX FT /label= OTHER
XX FT /note= "OTHER= H2NC(=NH)NH(CH2)CO. Beta-Ala (not
XX FT defined)"
XX PN WO9501371-Al.
XX PD 12-JAN-1995.
XX PF 22-JUN-1994; 94WO-JP000999.
XX PR 30-JUN-1993; 93JP-00186755.
XX PA (YAWA ) NIPPON STEEL CORP.
XX PI Sato Y, Hayashi Y, Katada J, Takiguchi Y;
XX DR WPI; 1995-060950/08.
XX PT New RGD peptide(s) useful as anti:platelet aggregation agents - contain
XX FT guanidino or amidino gp. at N-terminal to increase stability.
XX PS Disclosure; Page 10; 34pp; Japanese.
XX CC The invention describes peptides of amino acid sequence (I) and their
XX CC salts. (I) are useful as platelet aggregation inhibitors and are easily
XX CC absorbed by the body. Due to the presence of the N-terminal guanidino or
XX CC amidino group, (I) are provided with excellent stability so that their
XX CC activity can be exhibited for an effective time after administration.
XX CC Thereafter they are readily metabolised and expelled. This is the amino
XX CC acid sequence of a platelet aggregation inhibitor peptide.
XX SQ Sequence 1 AA;

QY 1 X 1
DB 1 R 1

RESULT 14
ADD94996
ID ADD94996 standard; peptide; 1 AA.
XX AC ADP94996;
XX DT 29-JAN-2004 (first entry)
XX DE Platelet aggregation inhibitor peptide #138.
XX KW platelet aggregation inhibitor; guanidino group; amidino group.
XX OS Unidentified.
XX FH Key Location/Qualifiers
XX FT Modified-site 1
XX FT /label= OTHER
XX FT /note= "OTHER= H2NC(=NH)NH(CH2)CO. Beta-Ala (not
XX FT defined)"
XX PN WO9501371-Al.
XX PD 12-JAN-1995.

QY 1 X 1
DB 1 R 1

RESULT 13
ADD94994
ID ADD94994 standard; peptide; 1 AA.
XX AC ADD94994;
XX DT 29-JAN-2004 (first entry)
XX DE Platelet aggregation inhibitor peptide #136.

```


PF 22-JUN-1994; 94WO-JP000999.
 XX
 PR 30-JUN-1993; 93JP-00186755.
 XX
 PA (YAWA) NIPPON STEEL CORP.
 XX
 PI Sato Y, Hayashi Y, Katada J, Takiguchi Y;
 XX
 DR WPI; 1995-060950/08.
 XX
 XX New RGD peptide(s) useful as anti-platelet aggregation agents - contain
 PT guanidino or amidino gp. at N-terminal to increase stability.
 XX
 XX Disclosure; Page 10; 34pp; Japanese.
 XX
 CC The invention describes peptides of amino acid sequence (I) and their
 CC salts. (I) are useful as platelet aggregation inhibitors and are easily
 CC absorbed by the body. Due to the presence of the N-terminal guanidino or
 CC amidino group, (I) are provided with excellent stability so that their
 CC activity can be exhibited for an effective time after administration.
 CC Thereafter they are readily metabolised and expelled. This is the amino
 CC acid sequence of a platelet aggregation inhibitor peptide.
 XX
 XX Sequence 1 AA;
 SQ

Query Match 0.0%; Score 0; DB 2; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
 DB 1 R 1

RESULT 16
 AAY46652
 ID AAY46652 standard; peptide; 1 AA.
 XX
 AC AAY46652;
 XX
 XX 01-DEC-1999 (first entry)
 DT
 XX
 DE Immunogenic peptide having a human leukocyte antigen binding motif #1263.
 XX
 KW Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
 KW immune response; T cell activation; major histocompatibility complex;
 KW cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
 KW prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
 KW vaccine; immunisation.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 XX WO9945954-A1.
 XX
 XX 16-SEP-1999.
 PD
 XX
 PF 13-MAR-1998; 98WO-US005039.
 XX
 PR 13-MAR-1998; 98WO-US005039.
 XX
 XX (EPIM-) EPIMUNE INC.
 PA
 XX
 PI Sette A, Kubo RT, Sidney J, Celis E, Grey HM, Southwood S;
 XX
 DR WPI; 1999-551214/46.
 XX
 PT New immunogenic peptides with HLA binding motif, useful in treatment and
 PT diagnosis of cancers and viral diseases.
 XX
 XX Claim 1; Page 80; 150pp; English.
 PS
 XX
 CC AAY45390 to AAY48214 represent specifically claimed immunogenic peptides
 CC having a human major histocompatibility complex (MHC) Class I (also known
 CC as human leukocyte antigen (HLA)) binding motif. The immunogenic peptides
 CC can bind to a specific HLA allele (i.e. HLA-A subtypes HLA-A2.1, A1, A3.2
 CC or A24.1 or HLA-B or C) and induce a cytotoxic T cell response against
 CC the antigen from which the peptide is derived. Cytotoxic T lymphocytes
 CC (CTLs) which destroy antigen-bearing cells are normally induced by an
 CC antigen in the form of a peptide fragment bound to a HLA molecule, rather
 CC than the intact foreign antigen itself, and are particularly important in
 CC tumour rejection and in fighting viral infections. The peptides are
 CC therefore useful therapeutically to treat or prevent viral infections and
 CC cancers in mammals (especially humans) e.g. prostate cancer, hepatitis B
 CC and C, AIDS, and renal carcinoma. They can be administered as vaccines to
 CC elicit an immune response in individuals susceptible or otherwise at risk
 CC of viral infection or cancer, or used to treat chronic or acute
 CC conditions. They are also useful diagnostically, and can be used to

PF 22-JUN-1994; 94WO-JP000999.
 XX
 PR 30-JUN-1993; 93JP-00186755.
 XX
 PA (YAWA) NIPPON STEEL CORP.
 XX
 PI Sato Y, Hayashi Y, Katada J, Takiguchi Y;
 XX
 DR WPI; 1995-060950/08.
 XX
 XX New RGD peptide(s) useful as anti-platelet aggregation agents - contain
 PT guanidino or amidino gp. at N-terminal to increase stability.
 XX
 XX Disclosure; Page 10; 34pp; Japanese.
 XX
 CC The invention describes peptides of amino acid sequence (I) and their
 CC salts. (I) are useful as platelet aggregation inhibitors and are easily
 CC absorbed by the body. Due to the presence of the N-terminal guanidino or
 CC amidino group, (I) are provided with excellent stability so that their
 CC activity can be exhibited for an effective time after administration.
 CC Thereafter they are readily metabolised and expelled. This is the amino
 CC acid sequence of a platelet aggregation inhibitor peptide.
 XX
 XX Sequence 1 AA;
 SQ

Query Match 0.0%; Score 0; DB 2; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
 DB 1 R 1

RESULT 15
 ADD95000
 ID ADD95000 standard; peptide; 1 AA.
 XX
 AC ADD95000;
 XX
 XX 29-JAN-2004 (first entry)
 DT
 XX
 DE Platelet aggregation inhibitor peptide #142.
 XX
 KW platelet aggregation inhibitor; guanidino group; amidino group.
 XX
 OS Unidentified.
 XX
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1
 FT /label= OTHER
 FT /note= "OTHER= H2NC(=NH)NH(CH2)CO. Beta-Ala (not
 FT defined)"
 XX
 XX WO9501371-A1.
 XX
 XX 12-JAN-1995.
 PD
 XX
 XX 22-JUN-1994; 94WO-JP000999.
 XX
 XX 30-JUN-1993; 93JP-00186755.
 XX
 XX (YAWA) NIPPON STEEL CORP.
 PA
 XX
 PI Sato Y, Hayashi Y, Katada J, Takiguchi Y;
 XX
 DR WPI; 1995-060950/08.
 XX
 XX New RGD peptide(s) useful as anti-platelet aggregation agents - contain
 PT guanidino or amidino gp. at N-terminal to increase stability.
 XX
 XX Disclosure; Page 10; 34pp; Japanese.
 PS
 XX

CC induce a cytotoxic T cell response, by contacting a cytotoxic T cell with
 CC the peptide e.g. to produce CTLs ex vivo for infusion back into a
 CC patient. The polynucleotides encoding the immunogenic peptides are also
 CC useful therapeutically and for immunisation as above

XX SQ Sequence 1 AA;

Query Match 0.0%; Score 0; DB 2; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
 DB 1 Y 1

RESULT 17

AAM97834
 ID AAM97834 standard; peptide; 1 AA.

XX AC

XX AC AAM97834;

XX DT 24-JAN-2002 (first entry)

XX DE Human peptide #1109 encoded by a SNP oligonucleotide.

XX KW Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;
 KW neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;
 KW amyloid protein; angiotensin; apoptosis related protein; cadherin;
 KW cyclin; polymerase; histones; kinases; colony stimulating factor;
 KW complement related protein; cytochrome; kinesin; cytokine; interferon;
 KW interleukin; G-protein coupled receptor; thioesterase; inflammation;
 KW multifactorial disease; autoimmune disease; infection;
 KW nervous system disease.

XX OS Homo sapiens.

XX PN WO200147944-A2.

XX PD 05-JUL-2001.

XX PF 28-DEC-2000; 2000WO-US035498.

XX PR 28-DEC-1999; 99US-0173419P.

XX PR 27-DEC-2000; 2000US-00173419.

XX PA (CURA-) CURAGEN CORP.

XX PI Shimkets RA, Leach M;

XX DR WPI; 2001-465210/50.

XX PT Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,
 PT oncogenes and histones, useful for diagnosing and treating, e.g. cancer,
 PT autoimmune diseases and infections.

XX PS Disclosure; Page 3911; 4143pp; English.

XX CC The present invention relates to oligonucleotides (see AAL26793-AAL34659)
 CC encoding polymorphic variants of proteins related to amylases, amyloid
 CC proteins, angiotensin, apoptosis related proteins, cadherin, cyclin,
 CC polymerase, oncogenes, histones, kinases, colony stimulating factors,
 CC complement related proteins, cytochromes, kinesins, cytokines,
 CC interferons, interleukins, G-protein coupled receptors and thioesterases.
 CC The present sequence is a peptide encoded by one such oligonucleotide.
 CC The oligonucleotides and the peptides encoded by them may be used in the
 CC prevention, diagnosis and treatment of diseases associated with
 CC inappropriate expression of the proteins listed above. Disorders that may
 CC be prevented, diagnosed and/or treated include multifactorial diseases
 CC with a genetic component, such as autoimmune diseases (e.g. rheumatoid
 CC arthritis, multiple sclerosis, diabetes, systemic lupus erythematosus
 CC and Grave's disease), inflammation, cancer (e.g. cancers of the bladder,
 CC brain, breast, colon and kidney, leukaemia), diseases of the nervous

CC system and an infection of pathogenic organisms

XX SQ Sequence 1 AA;

Query Match 0.0%; Score 0; DB 4; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
 DB 1 E 1

RESULT 18

AAM97974

ID AAM97974 standard; peptide; 1 AA.

XX AC

XX AC AAM97974;

XX DT 24-JAN-2002 (first entry)

XX DE Human peptide #1249 encoded by a SNP oligonucleotide.

XX KW Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;
 KW neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;
 KW amyloid protein; angiotensin; apoptosis related protein; cadherin;
 KW cyclin; polymerase; histones; kinases; colony stimulating factor;
 KW complement related protein; cytochrome; kinesin; cytokine; interferon;
 KW interleukin; G-protein coupled receptor; thioesterase; inflammation;
 KW multifactorial disease; autoimmune disease; infection;
 KW nervous system disease.

XX OS Homo sapiens.

XX PN WO200147944-A2.

XX PD 05-JUL-2001.

XX PF 28-DEC-2000; 2000WO-US035498.

XX PR 28-DEC-1999; 99US-0173419P.

XX PR 27-DEC-2000; 2000US-00173419.

XX PA (CURA-) CURAGEN CORP.

XX PI Shimkets RA, Leach M;

XX DR WPI; 2001-465210/50.

XX PT Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,
 PT oncogenes and histones, useful for diagnosing and treating, e.g. cancer,
 PT autoimmune diseases and infections.

XX PS Disclosure; Page 3941; 4143pp; English.

XX CC The present invention relates to oligonucleotides (see AAL26793-AAL34659)
 CC encoding polymorphic variants of proteins related to amylases, amyloid
 CC proteins, angiotensin, apoptosis related proteins, cadherin, cyclin,
 CC polymerase, oncogenes, histones, kinases, colony stimulating factors,
 CC complement related proteins, cytochromes, kinesins, cytokines,
 CC interferons, interleukins, G-protein coupled receptors and thioesterases.
 CC The present sequence is a peptide encoded by one such oligonucleotide.
 CC The oligonucleotides and the peptides encoded by them may be used in the
 CC prevention, diagnosis and treatment of diseases associated with
 CC inappropriate expression of the proteins listed above. Disorders that may
 CC be prevented, diagnosed and/or treated include multifactorial diseases
 CC with a genetic component, such as autoimmune diseases (e.g. rheumatoid
 CC arthritis, multiple sclerosis, diabetes, systemic lupus erythematosus
 CC and Grave's disease), inflammation, cancer (e.g. cancers of the bladder,
 CC brain, breast, colon and kidney, leukaemia), diseases of the nervous
 CC system and an infection of pathogenic organisms

XX SQ Sequence 1 AA;

Query Match 0.0%; Score 0; DB 4; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
 DB 1 L 1

RESULT 19
 AAM97643
 ID AAM97643 standard; peptide; 1 AA.
 XX
 AC AAM97643;
 XX
 DT 24-JAN-2002 (first entry)
 XX
 DE Human peptide #918 encoded by a SNP oligonucleotide.
 XX
 KW Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;
 KW neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;
 KW amyloid protein; angiotensin; apoptosis related protein; cadherin;
 KW cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;
 KW complement related protein; cytochrome; kinesin; cytokine; interferon;
 KW interleukin; G-protein coupled receptor; thioesterase; inflammation;
 KW multifactorial disease; autoimmune disease; infection;
 KW nervous system disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200147944-A2.
 XX
 PD 05-JUL-2001.
 XX
 PF 28-DEC-2000; 2000WO-US035498.
 XX
 PR 28-DEC-1999; 99US-0173419P.
 PR 27-DEC-2000; 2000US-00173419.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX
 DR WPI; 2001-465210/50.
 XX
 PT Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,
 PT oncogenes and histones, useful for diagnosing and treating, e.g. cancer,
 PT autoimmune diseases and infections.
 XX
 PS Disclosure; Page 3869; 4143pp; English.
 XX
 CC The present invention relates to oligonucleotides (see AAL26793-AAL34659)
 CC encoding polymorphic variants of proteins related to amylases, amyloid
 CC proteins, angiotensin, apoptosis related proteins, cadherin, cyclin,
 CC polymerase, oncogenes, histones, kinases, colony stimulating factors,
 CC complement related proteins, cytochromes, kinesins, cytokines,
 CC interferons, interleukins, G-protein coupled receptors and thioesterases.
 CC The present sequence is a peptide encoded by one such oligonucleotide.
 CC The oligonucleotides and the peptides encoded by them may be used in the
 CC prevention, diagnosis and treatment of diseases associated with
 CC inappropriate expression of the proteins listed above. Disorders that may
 CC be prevented, diagnosed and/or treated include multifactorial diseases
 CC with a genetic component, such as autoimmune diseases (e.g. rheumatoid
 CC arthritis, multiple sclerosis, diabetes, systemic lupus erythematosus
 CC and Grave's disease), inflammation, cancer (e.g. cancers of the bladder,
 CC brain, breast, colon and kidney, leukaemia), diseases of the nervous
 CC system and an infection of pathogenic organisms
 XX
 SQ Sequence 1 AA;
 Query Match 0.0%; Score 0; DB 4; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
 DB 1 C 1

RESULT 20
 AAM98447
 ID AAM98447 standard; peptide; 1 AA.
 XX
 AC AAM98447;
 XX
 DT 24-JAN-2002 (first entry)
 XX
 DE Human peptide #1722 encoded by a SNP oligonucleotide.
 XX
 KW Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;
 KW neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;
 KW amyloid protein; angiotensin; apoptosis related protein; cadherin;
 KW cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;
 KW complement related protein; cytochrome; kinesin; cytokine; interferon;
 KW interleukin; G-protein coupled receptor; thioesterase; inflammation;
 KW multifactorial disease; autoimmune disease; infection;
 KW nervous system disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200147944-A2.
 XX
 PD 05-JUL-2001.
 XX
 PF 28-DEC-2000; 2000WO-US035498.
 XX
 PR 28-DEC-1999; 99US-0173419P.
 PR 27-DEC-2000; 2000US-00173419.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX
 DR WPI; 2001-465210/50.
 XX
 PT Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,
 PT oncogenes and histones, useful for diagnosing and treating, e.g. cancer,
 PT autoimmune diseases and infections.
 XX
 PS Disclosure; Page 4045; 4143pp; English.
 XX
 CC The present invention relates to oligonucleotides (see AAL26793-AAL34659)
 CC encoding polymorphic variants of proteins related to amylases, amyloid
 CC proteins, angiotensin, apoptosis related proteins, cadherin, cyclin,
 CC polymerase, oncogenes, histones, kinases, colony stimulating factors,
 CC complement related proteins, cytochromes, kinesins, cytokines,
 CC interferons, interleukins, G-protein coupled receptors and thioesterases.
 CC The present sequence is a peptide encoded by one such oligonucleotide.
 CC The oligonucleotides and the peptides encoded by them may be used in the
 CC prevention, diagnosis and treatment of diseases associated with
 CC inappropriate expression of the proteins listed above. Disorders that may
 CC be prevented, diagnosed and/or treated include multifactorial diseases
 CC with a genetic component, such as autoimmune diseases (e.g. rheumatoid
 CC arthritis, multiple sclerosis, diabetes, systemic lupus erythematosus
 CC and Grave's disease), inflammation, cancer (e.g. cancers of the bladder,
 CC brain, breast, colon and kidney, leukaemia), diseases of the nervous
 CC system and an infection of pathogenic organisms
 XX
 SQ Sequence 1 AA;
 Query Match 0.0%; Score 0; DB 4; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1

Db		1 L 1	
RESULT 21			
AAM98354	ID AAM98354 standard; peptide; 1 AA.		
XX AC	AAM98354;		
XX DT	24-JAN-2002 (first entry)		
XX XX	Human peptide #1629 encoded by a SNP oligonucleotide.		
DE DE	Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic; neuroprotective; antimicrobial; gene therapy; vaccine; cancer; amyloid protein; angiotensin; apoptosis related protein; cadherin; cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor; complement related protein; cytochrome; kinesin; cytokine; interferon; interleukin; G-protein coupled receptor; thioesterase; inflammation; multifactorial disease; autoimmune disease; infection; nervous system disease.		
XX KW	Homo sapiens.		
XX OS			
XX XX			
PN PN	WO200147944-A2.		
XX PD	05-JUL-2001.		
XX PF	28-DEC-2000; 2000WO-US035498.		
XX PP			
XX PR	28-DEC-1999; 99US-0173419P.		
PR PR	27-DEC-2000; 2000US-00173419.		
XX XX	(CURA-) CURAGEN CORP.		
PA PA	Shimkets RA, Leach M;		
XX PI			
XX PT	WPI; 2001-465210/50.		
DR DR	Polymeric nucleic acids encoding e.g. amylases, cyclins, polymerases, oncogenes and histones, useful for diagnosing and treating, e.g. cancer, autoimmune diseases and infections.		
PT PT			
PT PT			
XX PS	Disclosure; Page 4025; 4143pp; English.		
XX CC	The present invention relates to oligonucleotides (see AAL26793-AAL34659) encoding polymorphic variants of proteins related to amylases, amyloid proteins, angiotensin, apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes, histones, kinases, colony stimulating factors, complement related proteins, cytochromes, kinesins, cytokines, interferons, interleukins, G-protein coupled receptors and thioesterases. The present sequence is a peptide encoded by one such oligonucleotide. The oligonucleotides and the peptides encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate expression of the proteins listed above. Disorders that may be prevented, diagnosed and/or treated include multifactorial diseases with a genetic component, such as autoimmune diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes, systemic lupus erythematosus and Grave's disease), inflammation, cancer (e.g. cancers of the bladder, brain, breast, colon and kidney, leukaemia), diseases of the nervous system and an infection of pathogenic organisms		
XX CC			
SQ SQ	Sequence 1 AA;		
	Query Match 0.0%; Score 0; DB 4; Length 1;		
	Best Local Similarity 0.0%; Pred. No. 0;		
	Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
Qy Qy	1 X 1		
Db Db	1 L 1		
RESULT 23			
AAM53329	ID AAM53329 standard; peptide; 1 AA.		
XX XX	AAM53329;		
AC AC			

XX 09-NOV-2001 (first entry)
 XX Human nonconservative amino acid changing SNP related peptide SEQ:7024.
 DE Human; single nucleotide polymorphism; SNP; genome; gene therapy;
 KW protein therapy; vaccine; probe; diagnostic assay; detection;
 KW quantitation; restorative therapy; polymorphic.
 XX
 OS Homo sapiens.
 XX WO200140521-A2.
 PN 07-JUN-2001.
 PD 30-NOV-2000; 2000WO-US032758.
 XX 30-NOV-1999; 99US-0168138P.
 PF 29-NOV-2000; 2000US-00726173.
 XX (CURA-) CURAGEN CORP.
 PA Shimkets RA, Leach M;
 XX WPI; 2001-356160/37.
 PI Polymorphic nucleic acid sequences, useful in genetic testing and therapy.
 XX Claim 29; Page 2653; 2653pp; English.
 PS AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide sequences (I), which contain single nucleotide polymorphisms (SNPs).
 CC AAM53114 to AAM53329 represent peptides related to human polymorphic polynucleotide sequences. The sequences can be used in gene and protein therapy, and in vaccine production. (I) and the polypeptides encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate expression of polymorphic polypeptides. For example, (I) may be used to treat disorders by rectifying mutations or deletions in a patient's genome that affect the activity of polypeptides by expressing inactive proteins or to supplement the patients own production of polypeptide. Additionally, (I) and its complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. The polypeptides encoded by (I) may be used as antigens in the production of antibodies specific for polymorphic polypeptides. The antibodies may also be used to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of polymorphic polypeptides in samples
 XX Sequence 1 AA;
 SQ
 Query Match 0.0%; Score 0; DB 4; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 1 X 1
 Db 1 V 1
 RESULT 24
 AAM53291
 ID AAM53291 standard; peptide; 1 AA.
 XX
 AC AAM53291;
 XX
 DT 09-NOV-2001 (first entry)
 DE Human nonconservative amino acid changing SNP related peptide SEQ:6986.
 XX Human; single nucleotide polymorphism; SNP; genome; gene therapy;
 KW

KW protein therapy; vaccine; probe; diagnostic assay; detection;
 KW quantitation; restorative therapy; polymorphic.
 XX
 OS Homo sapiens.
 XX WO200140521-A2.
 PN 07-JUN-2001.
 PD 30-NOV-2000; 2000WO-US032758.
 XX 30-NOV-1999; 99US-0168138P.
 PF 29-NOV-2000; 2000US-00726173.
 XX (CURA-) CURAGEN CORP.
 PA Shimkets RA, Leach M;
 XX WPI; 2001-356160/37.
 PI Polymorphic nucleic acid sequences, useful in genetic testing and therapy.
 XX Claim 29; Page 2641; 2653pp; English.
 PS AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide sequences (I), which contain single nucleotide polymorphisms (SNPs).
 CC AAM53114 to AAM53329 represent peptides related to human polymorphic polynucleotide sequences. The sequences can be used in gene and protein therapy, and in vaccine production. (I) and the polypeptides encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate expression of polymorphic polypeptides. For example, (I) may be used to treat disorders by rectifying mutations or deletions in a patient's genome that affect the activity of polypeptides by expressing inactive proteins or to supplement the patients own production of polypeptide. Additionally, (I) and its complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. The polypeptides encoded by (I) may be used as antigens in the production of antibodies specific for polymorphic polypeptides. The antibodies may also be used to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of polymorphic polypeptides in samples
 XX Sequence 1 AA;
 SQ
 Query Match 0.0%; Score 0; DB 4; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 1 X 1
 Db 1 L 1
 RESULT 25
 AAM53328
 ID AAM53328 standard; peptide; 1 AA.
 XX
 AC AAM53328;
 XX
 DT 09-NOV-2001 (first entry)
 DE Human nonconservative amino acid changing SNP related peptide SEQ:7023.
 XX Human; single nucleotide polymorphism; SNP; genome; gene therapy;
 KW protein therapy; vaccine; probe; diagnostic assay; detection;
 KW quantitation; restorative therapy; polymorphic.
 XX
 OS Homo sapiens.
 XX WO200140521-A2.
 PN

DR WPI; 2001-356160/37.
XX Polymorphic nucleic acid sequences, useful in genetic testing and
PT therapy.
XX
XX Claim 29; Page 2619; 2653pp; English.
XX
XX AA173060 to AA173867 represent isolated human polymorphic polynucleotide
CC sequences (I), which contain single nucleotide polymorphisms (SNPs).
CC AAM53114 to AAM53329 represent peptides related to human polymorphic
CC polynucleotide sequences. The sequences can be used in gene and protein
CC therapy, and in vaccine production. (I) and the polypeptides encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate expression of polymorphic polypeptides. For
CC example, (I) may be used to treat disorders by rectifying mutations or
CC deletions in a patient's genome that affect the activity of polypeptides
CC by expressing inactive proteins or to supplement the patients own
CC production of polypeptide. Additionally, (I) and its complementary
CC sequences may also be used as DNA probes in diagnostic assays to detect
CC and quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. The
CC polypeptides encoded by (I) may be used as antigens in the production of
CC antibodies specific for polymorphic polypeptides. The antibodies may also
CC be used to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of polymorphic
CC polypeptides in samples
XX
XX Sequence 1 AA;
Query Match 0.0%; Score 0; DB 4; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 X 1
Db 1 H 1
RESULT 28
AAB91029
ID AAB91029 standard; peptide; 1 AA.
XX
XX AAB91029;
XX
XX 22-JUN-2001 (first entry)
XX
XX Thyrotropin releasing hormone (THR) related peptide SEQ ID NO:203.
XX
XX Protection; endogenous therapeutic peptide; peptidase; conjugation;
XX blood component; modification; succinimidyl; maleimido group; amino;
XX hydroxyl; thiol; hormone; growth factor; neurotransmitter.
XX
XX Homo sapiens.
XX Synthetic.
XX
XX WO200069900-A2.
XX
XX 23-NOV-2000.
XX
XX 17-MAY-2000; 2000WO-US013576.
XX
XX 17-MAY-1999; 99US-0134406P.
XX 10-SEP-1999; 99US-0153406P.
XX 15-OCT-1999; 99US-0159783P.
XX
XX (CONJ-) CONJUCHEM INC.
XX
XX Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudeau K;
XX WPI; 2001-112059/12.
XX
XX Modifying and attaching therapeutic peptides to albumin prevents
PT peptidase degradation, useful for increasing length of in vivo activity.
XX

XX Disclosure; Page 256; 733pp; English.
XX
XX The present invention describes a modified therapeutic peptide (I)
CC comprising a therapeutically active amino acid region (III) and a
CC reactive group (II) (e.g. succinimidyl and maleimido groups) attached to
CC a less therapeutically active amino acid region (IV), which covalently
CC bonds with amino/hydroxyl/thiol groups on blood components to form a
CC peptidase stabilised therapeutic peptide composed of 3-50 amino acids.
CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth
CC factors and neurotransmitters, to protect them from peptidase activity in
CC vivo for the treatment of various disorders. Endogenous therapeutic
CC peptides are not suitable as drug candidates as they require frequent
CC administration due to rapid degradation by peptidases in the body.
CC Modifying and attaching therapeutic peptides to albumin prevents or
CC reduces the action of peptidases to increase length of activity (half
CC life) and specificity as bonding to large molecules decreases
CC intracellular uptake and interference with physiological processes.
CC AAB90829 to AAB92441 represent peptides which can be used in the
CC exemplification of the present invention
XX
XX Sequence 1 AA;
Query Match 0.0%; Score 0; DB 4; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 X 1
Db 1 T 1
RESULT 29
AAB91739
ID AAB91739 standard; peptide; 1 AA.
XX
XX AAB91739;
XX
XX 22-JUN-2001 (first entry)
XX
XX Opioid peptide SEQ ID NO:915.
XX
XX Protection; endogenous therapeutic peptide; peptidase; conjugation;
XX blood component; modification; succinimidyl; maleimido group; amino;
XX hydroxyl; thiol; hormone; growth factor; neurotransmitter.
XX
XX Homo sapiens.
XX Synthetic.
XX
XX WO200069900-A2.
XX
XX 23-NOV-2000.
XX
XX 17-MAY-2000; 2000WO-US013576.
XX
XX 17-MAY-1999; 99US-0134406P.
XX 10-SEP-1999; 99US-0153406P.
XX 15-OCT-1999; 99US-0159783P.
XX
XX (CONJ-) CONJUCHEM INC.
XX
XX Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudeau K;
XX WPI; 2001-112059/12.
XX
XX Modifying and attaching therapeutic peptides to albumin prevents
PT peptidase degradation, useful for increasing length of in vivo activity.
XX
XX Disclosure; Page 492; 733pp; English.
XX
XX The present invention describes a modified therapeutic peptide (I)
CC comprising a therapeutically active amino acid region (III) and a
CC reactive group (II) (e.g. succinimidyl and maleimido groups) attached to

CC a less therapeutically active amino acid region (IV), which covalently
 CC bonds with amino/hydroxyl/thiol groups on blood components to form a
 CC peptidase stabilised therapeutic peptide composed of 3-50 amino acids.
 CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth
 CC factors and neurotransmitters, to protect them from peptidase activity in
 CC vivo for the treatment of various disorders. Endogenous therapeutic
 CC peptides are not suitable as drug candidates as they require frequent
 CC administration due to rapid degradation by peptidases in the body.
 CC Modifying and attaching therapeutic peptides to albumin prevents or
 CC reduces the action of peptidases to increase length of activity (half
 CC life) and specificity as bonding to large molecules decreases
 CC intracellular uptake and interference with physiological processes.
 CC AAB90829 to AAB92441 represent peptides which can be used in the
 CC exemplification of the present invention
 XX Sequence 1 AA;
 SQ

Query Match 0.0%; Score 0; DB 4; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1

Db 1 Y 1

RESULT 30

AAB92150

ID AAB92150 standard; peptide; 1 AA.

XX AAB92150;

XX 22-JUN-2001 (first entry)

XX Polypeptide SEQ ID NO:1326.

XX Protection; endogenous therapeutic peptide; peptidase; conjugation;
 KW blood component; modification; succinimidyl; maleimido group; amino;
 KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.
 XX

OS Homo sapiens.

OS Synthetic.

FN WO200069900-A2.

XX 23-NOV-2000.

XX 17-MAY-2000; 2000WO-US013576.

XX 17-MAY-1999; 99US-0134406P.

PR 10-SEP-1999; 99US-0153406P.

PR 15-OCT-1999; 99US-0159783P.

XX (CONJ-) CONJUCHEM INC.

XX Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudeau K;

PI WPI; 2001-112059/12.

XX Modifying and attaching therapeutic peptides to albumin prevents

XX peptidase degradation, useful for increasing length of in vivo activity.

XX Disclosure; Page 630; 733pp; English.

XX The present invention describes a modified therapeutic peptide (I)

XX comprising a therapeutically active amino acid region (III) and a

XX reactive group (II) (e.g. succinimidyl and maleimido groups) attached to

XX bonds with amino/hydroxyl/thiol groups on blood components to form a

XX peptidase stabilised therapeutic peptide composed of 3-50 amino acids.

XX (I) are useful for modifying therapeutic peptides e.g. hormones, growth

XX factors and neurotransmitters, to protect them from peptidase activity in

XX vivo for the treatment of various disorders. Endogenous therapeutic

CC peptides are not suitable as drug candidates as they require frequent
 CC administration due to rapid degradation by peptidases in the body.
 CC Modifying and attaching therapeutic peptides to albumin prevents or
 CC reduces the action of peptidases to increase length of activity (half
 CC life) and specificity as bonding to large molecules decreases
 CC intracellular uptake and interference with physiological processes.
 CC AAB90829 to AAB92441 represent peptides which can be used in the
 CC exemplification of the present invention
 XX Sequence 1 AA;
 SQ

Query Match 0.0%; Score 0; DB 4; Length 1;

Best Local Similarity 0.0%; Pred. No. 0;

Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1

Db 1 K 1

RESULT 31

AAB91892

ID AAB91892 standard; peptide; 1 AA.

XX AAB91892;

XX 22-JUN-2001 (first entry)

XX Apoptosis related peptide SEQ ID NO:1068.

XX Protection; endogenous therapeutic peptide; peptidase; conjugation;
 KW blood component; modification; succinimidyl; maleimido group; amino;
 KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.
 XX

OS Homo sapiens.

OS Synthetic.

FN WO200069900-A2.

XX 23-NOV-2000.

XX 17-MAY-2000; 2000WO-US013576.

XX 17-MAY-1999; 99US-0134406P.

PR 10-SEP-1999; 99US-0153406P.

PR 15-OCT-1999; 99US-0159783P.

XX (CONJ-) CONJUCHEM INC.

XX Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudeau K;

PI WPI; 2001-112059/12.

XX Modifying and attaching therapeutic peptides to albumin prevents

XX peptidase degradation, useful for increasing length of in vivo activity.

XX Disclosure; Page 545; 733pp; English.

XX The present invention describes a modified therapeutic peptide (I)

XX comprising a therapeutically active amino acid region (III) and a

XX reactive group (II) (e.g. succinimidyl and maleimido groups) attached to

XX bonds with amino/hydroxyl/thiol groups on blood components to form a

XX peptidase stabilised therapeutic peptide composed of 3-50 amino acids.

XX (I) are useful for modifying therapeutic peptides e.g. hormones, growth

XX factors and neurotransmitters, to protect them from peptidase activity in

XX vivo for the treatment of various disorders. Endogenous therapeutic

CC AAB90829 to AAB92441 represent peptides which can be used in the
 CC exemplification of the present invention

XX Sequence 1 AA;

Query Match 0.0%; Score 0; DB 4; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
 DB 1 D 1

RESULT 32
 AAB91546
 ID AAB91546 standard; peptide; 1 AA.

AC AAB91546;

XX 22-JUN-2001 (first entry)

DE Endothelins and related peptides SEQ ID NO:722.

KW Protection; endogenous therapeutic peptide; peptidase; conjugation;
 KW blood component; modification; succinimidyl; maleimido group; amino;
 KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.

XX Homo sapiens.
 OS Synthetic.

XX WO2000069900-A2.

PN 23-NOV-2000.

XX 17-MAY-2000; 2000WO-US013576.

PR 17-MAY-1999; 99US-0134406P.

PR 10-SEP-1999; 99US-0153406P.

PR 15-OCT-1999; 99US-0159783P.

XX (CONJ-) CONJUCHEM INC.

XX Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudeau K;

XX WPI; 2001-112059/12.

XX Modifying and attaching therapeutic peptides to albumin prevents
 PT peptidase degradation, useful for increasing length of in vivo activity.

PS Disclosure; Page 433; 733pp; English.

CC The present invention describes a modified therapeutic peptide (I)
 CC comprising a therapeutically active amino acid region (III) and a
 CC reactive group (II) (e.g. succinimidyl and maleimido groups) attached to
 CC a less therapeutically active amino acid region (IV), which covalently
 CC bonds with amino/hydroxyl/thiol groups on blood components to form a
 CC peptidase stabilised therapeutic peptide composed of 3-50 amino acids.
 CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth
 CC factors and neurotransmitters, to protect them from peptidase activity in
 CC vivo for the treatment of various disorders. Endogenous therapeutic
 CC peptides are not suitable as drug candidates as they require frequent
 CC administration due to rapid degradation by peptidases in the body.
 CC Modifying and attaching therapeutic peptides to albumin prevents or
 CC reduces the action of peptidases to increase length of activity (half
 CC life) and specificity as bonding to large molecules decreases
 CC intracellular uptake and interference with physiological processes.
 CC AAB90829 to AAB92441 represent peptides which can be used in the
 CC exemplification of the present invention

XX Sequence 1 AA;

Query Match 0.0%; Score 0; DB 4; Length 1;

Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
 DB 1 R 1

RESULT 33
 AAB92392
 ID AAB92392 standard; peptide; 1 AA.

XX AAB92392;

XX 22-JUN-2001 (first entry)

DE Miscellaneous peptide SEQ ID NO:1568.

KW Protection; endogenous therapeutic peptide; peptidase; conjugation;
 KW blood component; modification; succinimidyl; maleimido group; amino;
 KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.

XX Homo sapiens.
 OS Synthetic.

XX WO2000069900-A2.

XX 23-NOV-2000.

XX 17-MAY-2000; 2000WO-US013576.

PR 17-MAY-1999; 99US-0134406P.

PR 10-SEP-1999; 99US-0153406P.

PR 15-OCT-1999; 99US-0159783P.

XX (CONJ-) CONJUCHEM INC.

XX Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudeau K;

XX WPI; 2001-112059/12.

XX Modifying and attaching therapeutic peptides to albumin prevents

PT peptidase degradation, useful for increasing length of in vivo activity.

PS Disclosure; Page 717; 733pp; English.

CC The present invention describes a modified therapeutic peptide (I)
 CC comprising a therapeutically active amino acid region (III) and a
 CC reactive group (II) (e.g. succinimidyl and maleimido groups) attached to
 CC a less therapeutically active amino acid region (IV), which covalently
 CC bonds with amino/hydroxyl/thiol groups on blood components to form a
 CC peptidase stabilised therapeutic peptide composed of 3-50 amino acids.
 CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth
 CC factors and neurotransmitters, to protect them from peptidase activity in
 CC vivo for the treatment of various disorders. Endogenous therapeutic
 CC peptides are not suitable as drug candidates as they require frequent
 CC administration due to rapid degradation by peptidases in the body.
 CC Modifying and attaching therapeutic peptides to albumin prevents or
 CC reduces the action of peptidases to increase length of activity (half
 CC life) and specificity as bonding to large molecules decreases
 CC intracellular uptake and interference with physiological processes.
 CC AAB90829 to AAB92441 represent peptides which can be used in the
 CC exemplification of the present invention

XX Sequence 1 AA;

Query Match 0.0%; Score 0; DB 4; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
 DB 1 L 1


```

RESULT 34
AAB91665
ID AAB91665 standard; peptide; 1 AA.
XX
AC AAB91665;
XX
DT 22-JUN-2001 (first entry)
XX
DE Opioid peptide SEQ ID NO:841.
XX
KW Protection; endogenous therapeutic peptide; peptidase; conjugation;
KW blood component; modification; succinimidyl; maleimido group; amino;
KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200069900-A2.
XX
PD 23-NOV-2000.
XX
PF 17-MAY-2000; 2000WO-US013576.
XX
PR 17-MAY-1999; 99US-0134406P.
PR 10-SEP-1999; 99US-0153406P.
PR 15-OCT-1999; 99US-0159783P.
XX
XX (CONJ-) CONJUCHEM INC.
XX
XX Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudeau K;
PI WPI; 2001-112059/12.
XX
XX Modifying and attaching therapeutic peptides to albumin prevents
PT peptidase degradation, useful for increasing length of in vivo activity.
XX
PS Disclosure; Page 470; 733pp; English.
XX
XX The present invention describes a modified therapeutic peptide (I)
CC comprising a therapeutically active amino acid region (III) and a
CC reactive group (II) (e.g. succinimidyl and maleimido groups) attached to
CC a less therapeutically active amino acid region (IV), which covalently
CC bonds with amino/hydroxyl/thiol groups on blood components to form a
CC peptidase stabilised therapeutic peptide composed of 3-50 amino acids.
CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth
CC factors and neurotransmitters, to protect them from peptidase activity in
CC vivo for the treatment of various disorders. Endogenous therapeutic
CC peptides are not suitable as drug candidates as they require frequent
CC administration due to rapid degradation by peptidases in the body.
CC Modifying and attaching therapeutic peptides to albumin prevents or
CC reduces the action of peptidases to increase length of activity (half
CC life) and specificity as bonding to large molecules decreases
CC intracellular uptake and interference with physiological processes.
CC AAB90829 to AAB92441 represent peptides which can be used in the
CC exemplification of the present invention
XX
XX Sequence 1 AA;
SQ
Query Match 0.0%; Score 0; DB 4; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 X 1
Db 1 C 1
RESULT 35
AAG99966
ID AAG99966 standard; protein; 1 AA.
XX
AC AAG99966;
XX
DT 11-SEP-2003 (revised)
DT 27-SEP-2001 (first entry)
XX
DE ERA binding domain polypeptide SEQ ID NO 408.
XX
KW ERA binding domain; Escherichia coli; GTPase; antimicrobial;
KW antibacterial; antibiotic; pathogenesis; infection; vaccine;
KW peptide therapy.
XX
OS Escherichia coli.
XX
PN WO200153458-A2.
XX
PD 26-JUL-2001.
XX
PF 17-JAN-2001; 2001WO-US001786.
XX
PR 18-JAN-2000; 2000US-0176870P.
XX
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PA (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI Lupas AN, Pearce KH;
XX
XX WPI; 2001-476108/51.
XX
XX New ERA binding domain polypeptides and polynucleotides encoding them,
XX useful as research reagents and materials for discovery of treatments and
XX diagnostics for diseases, or for genetic immunization.
XX
XX Claim 1; Page 267; 279pp; English.
XX
XX The present invention relates to ERA binding domain polypeptides
CC (AAG99559-AAG99989 and AAM00010-AAM00021). The era gene in Escherichia
CC coli codes for an essential GTPase protein able to autophosphorylate at
CC serine and/or threonine residues. The protein has potential antimicrobial
CC and antibacterial activity and is useful in screening for antagonists,
CC agonists and for compounds with antibiotic activity. The proteins are
CC also useful in determining their role in pathogenesis of infection,
CC dysfunction and disease and could be used as part of a vaccine and/or
CC peptide therapy. (Updated on 11-SEP-2003 to standardise OS field)
XX
XX Sequence 1 AA;
SQ
Query Match 0.0%; Score 0; DB 4; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 X 1
Db 1 C 1
RESULT 36
AAM00011
ID AAM00011 standard; protein; 1 AA.
XX
AC AAM00011;
XX
DT 11-SEP-2003 (revised)
DT 27-SEP-2001 (first entry)
XX
DE ERA binding domain polypeptide SEQ ID NO 433.
XX
KW ERA binding domain; Escherichia coli; GTPase; antimicrobial;
KW antibacterial; antibiotic; pathogenesis; infection; vaccine;
KW peptide therapy.
XX
OS Escherichia coli.
XX
PN WO200153458-A2.

```


New ERA binding domain polypeptides and polynucleotides encoding them,
useful as research reagents and materials for discovery of treatments and
diagnostics for diseases, or for genetic immunization.

Claim 1; Page 272; 279pp; English.

The present invention relates to ERA binding domain polypeptides
(AAG99559-AAG99989 and AAM00010-AAM00021). The era gene in Escherichia
coli codes for an essential GTPase protein able to autophosphorylate at
serine and/or threonine residues. The protein has potential antimicrobial
and antibacterial activity and is useful in screening for antagonists,
agonists and for compounds with antibiotic activity. The proteins are
also useful in determining their role in pathogenesis of infection,
dysfunction and disease and could be used as part of a vaccine and/or
cystic therapy. (Updated on 11-SEP-2003 to standardise OS field)

Sequence 1 AA;

Query Match 0.0%; Score 0; DB 4; Length 1;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy 1 X 1
 |
Db 1 X 1

RESULT 38

AAG99983

ID AAG99983 standard; protein; 1 AA.

XX AAG99983;

XX AC

XX XX

DT 11-SEP-2003 (revised)

DT 27-SEP-2001 (first entry)

XX ERA binding domain polypeptide SEQ ID NO 425.

XX ERA binding domain; Escherichia coli; GTPase; antimicrobial;
KW antibacterial; antibiotic; pathogenesis; infection; vaccine;
KW peptide therapy.

XX Escherichia coli.

OS Escherichia coli.

XX WO200153458-A2.

FN 26-JUL-2001.

PD 17-JAN-2001; 2001WO-US001786.

XX 18-JAN-2000; 2000US-0176870P.

PF (SMIK) SMITHKLINE BEECHAM CORP.
PR (SMIK) SMITHKLINE BEECHAM PLC.

XX Lupas AN, Pearce KH;

PI WPI; 2001-476108/51.

XX New ERA binding domain polypeptides and polynucleotides encoding them,
PT useful as research reagents and materials for discovery of treatments and
FT diagnostics for diseases, or for genetic immunization.

PS Claim 1; Page 271; 279pp; English.

XX The present invention relates to ERA binding domain polypeptides
CC (AAG99559-AAG99989 and AAM00010-AAM00021). The era gene in Escherichia
CC coli codes for an essential GTPase protein able to autophosphorylate at
CC serine and/or threonine residues. The protein has potential antimicrobial
CC and antibacterial activity and is useful in screening for antagonists,
CC agonists and for compounds with antibiotic activity. The proteins are
CC also useful in determining their role in pathogenesis of infection,
CC dysfunction and disease and could be used as part of a vaccine and/or
CC cystic therapy. (Updated on 11-SEP-2003 to standardise OS field)


```

CC peptide therapy. (Updated on 11-SEP-2003 to standardise OS field)
XX
SQ Sequence 1 AA;

Query Match      0.0%; Score 0; DB 4; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 X 1
Db      1 C 1

RESULT 39
AAM00013
ID AAM00013 standard; protein; 1 AA.
XX
AC AAM00013;
XX
DT 11-SEP-2003 (revised)
DT 27-SEP-2001 (first entry)
XX
DE ERA binding domain polypeptide SEQ ID NO 435.
XX
KW ERA binding domain; Escherichia coli; GTPase; antimicrobial;
KW antibacterial; antibiotic; pathogenesis; infection; vaccine;
KW peptide therapy.
XX
OS Escherichia coli.
XX
PN WO200153458-A2.
XX
PD 26-JUL-2001.
XX
PF 17-JAN-2001; 2001WO-US001786.
XX
PR 18-JAN-2000; 2000US-0176870P.
XX
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PA (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI Lupas AN, Pearce KH;
XX
WPI; 2001-476108/51.
XX
PT New ERA binding domain polypeptides and polynucleotides encoding them,
PT useful as research reagents and materials for discovery of treatments and
PT diagnostics for diseases, or for genetic immunization.
XX
PS Claim 1; Page 274; 279pp; English.
XX
CC The present invention relates to ERA binding domain polypeptides
CC (AAG99559-AAG99989 and AAM00010-AAM00021). The era gene in Escherichia
CC coli codes for an essential GTPase protein able to autophosphorylate at
CC serine and/or threonine residues. The protein has potential antimicrobial
CC and antibacterial activity and is useful in screening for antagonists,
CC agonists and for compounds with antibiotic activity. The proteins are
CC also useful in determining their role in pathogenesis of infection,
CC dysfunction and disease and could be used as part of a vaccine and/or
CC peptide therapy. (Updated on 11-SEP-2003 to standardise OS field)
XX
SQ Sequence 1 AA;

Query Match      0.0%; Score 0; DB 4; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 X 1
Db      1 Q 1

RESULT 40

```

```

AAM00016
ID AAM00016 standard; protein; 1 AA.
XX
AC AAM00016;
XX
DT 11-SEP-2003 (revised)
DT 27-SEP-2001 (first entry)
XX
DE ERA binding domain polypeptide SEQ ID NO 438.
XX
KW ERA binding domain; Escherichia coli; GTPase; antimicrobial;
KW antibacterial; antibiotic; pathogenesis; infection; vaccine;
KW peptide therapy.
XX
OS Escherichia coli.
XX
PN WO200153458-A2.
XX
PD 26-JUL-2001.
XX
PF 17-JAN-2001; 2001WO-US001786.
XX
PR 18-JAN-2000; 2000US-0176870P.
XX
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PA (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI Lupas AN, Pearce KH;
XX
WPI; 2001-476108/51.
XX
PT New ERA binding domain polypeptides and polynucleotides encoding them,
PT useful as research reagents and materials for discovery of treatments and
PT diagnostics for diseases, or for genetic immunization.
XX
PS Claim 1; Page 275; 279pp; English.
XX
CC The present invention relates to ERA binding domain polypeptides
CC (AAG99559-AAG99989 and AAM00010-AAM00021). The era gene in Escherichia
CC coli codes for an essential GTPase protein able to autophosphorylate at
CC serine and/or threonine residues. The protein has potential antimicrobial
CC and antibacterial activity and is useful in screening for antagonists,
CC agonists and for compounds with antibiotic activity. The proteins are
CC also useful in determining their role in pathogenesis of infection,
CC dysfunction and disease and could be used as part of a vaccine and/or
CC peptide therapy. (Updated on 11-SEP-2003 to standardise OS field)
XX
SQ Sequence 1 AA;

Query Match      0.0%; Score 0; DB 4; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 X 1
Db      1 C 1

Search completed: April 13, 2005, 17:11:30
Job time : 12.3345 secs

```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2005, 17:03:40 ; Search time 2.86149 Seconds
(without alignments)
286.963 Million cell updates/sec

Title: US-09-998-861-14
Perfect score: 11
Sequence: 1 XXXXXXXXXX 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pap:*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pap:*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pap:*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pap:*
5: /cgn2_6/ptodata/1/iaa/PTUS COMB.pap:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	0	0.0	1	1 US-07-820-154A-12	Sequence 12, Appl
2	0	0.0	1	1 US-07-791-213D-24	Sequence 24, Appl
3	0	0.0	1	1 US-07-791-213D-40	Sequence 40, Appl
4	0	0.0	1	1 US-08-174-365A-57	Sequence 57, Appl
5	0	0.0	1	1 US-07-789-913-23	Sequence 23, Appl
6	0	0.0	1	1 US-07-789-913-25	Sequence 25, Appl
7	0	0.0	1	1 US-07-789-913-25	Sequence 25, Appl
8	0	0.0	1	1 US-08-049-794-23	Sequence 23, Appl
9	0	0.0	1	1 US-08-049-794-25	Sequence 25, Appl
10	0	0.0	1	1 US-08-433-037-12	Sequence 12, Appl
11	0	0.0	1	1 US-08-448-606-4	Sequence 4, Appl
12	0	0.0	1	1 US-07-869-933-16	Sequence 16, Appl
13	0	0.0	1	1 US-08-293-150A-24	Sequence 24, Appl
14	0	0.0	1	1 US-08-293-150A-40	Sequence 40, Appl
15	0	0.0	1	1 US-08-496-847-23	Sequence 23, Appl
16	0	0.0	1	1 US-08-496-847-25	Sequence 25, Appl
17	0	0.0	1	2 US-08-742-774-23	Sequence 23, Appl
18	0	0.0	1	2 US-08-742-774-25	Sequence 25, Appl
19	0	0.0	1	2 US-08-675-354-23	Sequence 23, Appl
20	0	0.0	1	2 US-08-675-354-25	Sequence 25, Appl
21	0	0.0	1	2 US-08-097-554A-12	Sequence 12, Appl
22	0	0.0	1	2 US-08-965-918-23	Sequence 23, Appl
23	0	0.0	1	2 US-08-965-918-25	Sequence 25, Appl
24	0	0.0	1	2 US-09-138-439-23	Sequence 23, Appl
25	0	0.0	1	2 US-09-138-439-25	Sequence 25, Appl
26	0	0.0	1	3 US-08-480-640A-12	Sequence 12, Appl
27	0	0.0	1	3 US-08-613-400A-23	Sequence 23, Appl
28	0	0.0	1	3 US-08-613-400A-25	Sequence 25, Appl

28	0	0.0	1	3 US-08-801-092-10	Sequence 10, Appl
29	0	0.0	1	3 US-08-801-092-17	Sequence 17, Appl
30	0	0.0	1	3 US-08-801-092-24	Sequence 24, Appl
31	0	0.0	1	3 US-08-801-092-31	Sequence 31, Appl
32	0	0.0	1	3 US-08-801-092-38	Sequence 38, Appl
33	0	0.0	1	3 US-08-801-092-45	Sequence 45, Appl
34	0	0.0	1	3 US-09-298-017-23	Sequence 23, Appl
35	0	0.0	1	3 US-09-298-017-25	Sequence 25, Appl
36	0	0.0	1	3 US-08-295-802-12	Sequence 12, Appl
37	0	0.0	1	3 US-09-392-979A-23	Sequence 23, Appl
38	0	0.0	1	3 US-09-392-979A-25	Sequence 25, Appl
39	0	0.0	1	3 US-09-103-663-16	Sequence 16, Appl
40	0	0.0	1	3 US-08-488-237A-12	Sequence 12, Appl
41	0	0.0	1	3 US-09-117-927-5	Sequence 5, Appl
42	0	0.0	1	3 US-08-375-992A-12	Sequence 12, Appl
43	0	0.0	1	3 US-09-315-113-10	Sequence 10, Appl
44	0	0.0	1	3 US-09-315-113-17	Sequence 17, Appl
45	0	0.0	1	3 US-09-315-113-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-07-820-154A-12
; Sequence 12, Application US/07820154A
; Patent No. 5382425
; GENERAL INFORMATION:
; APPLICANT: Cochran Ph.D., Mark D
; APPLICANT: Junker M.S., David E
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/820,154A
; FILING DATE: 19920113
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-820-154A-12

Query Match 0.0%; Score 0; DB 1; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 1 V 1

RESULT 2
US-07-791-213D-24


```
; Sequence 24, Application US/07791213D
; Patent No. 5409895
; GENERAL INFORMATION:
; APPLICANT: MORISHITA, Hideaki
; APPLICANT: KANAMORI, Toshinori
; APPLICANT: NOBUHARA, Masahiro
; TITLE OF INVENTION: POLYPEPTIDE, DNA FRAGMENT ENCODING THE
; TITLE OF INVENTION: SAME AND PROCESS FOR PRODUCING THE SAME, AND ENZYME
; TITLE OF INVENTION: INHIBITION PROCESS, DRUG COMPOSITION AND METHODS OF
; TITLE OF INVENTION: TREATING USING THE SAME
; NUMBER OF SEQUENCES: 108
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/791,213D
; FILING DATE: 13-NOV-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-306745
; FILING DATE: 13-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 029650-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-791-213D-24

Query Match 0.0%; Score 0; DB 1; Length 1;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 C 1

RESULT 3
US-07-791-213D-40
; Sequence 40, Application US/07791213D
; Patent No. 5409895
; GENERAL INFORMATION:
; APPLICANT: MORISHITA, Hideaki
; APPLICANT: KANAMORI, Toshinori
; APPLICANT: NOBUHARA, Masahiro
; TITLE OF INVENTION: POLYPEPTIDE, DNA FRAGMENT ENCODING THE
; TITLE OF INVENTION: SAME AND PROCESS FOR PRODUCING THE SAME, AND ENZYME
; TITLE OF INVENTION: INHIBITION PROCESS, DRUG COMPOSITION AND METHODS OF
; TITLE OF INVENTION: TREATING USING THE SAME
; NUMBER OF SEQUENCES: 108
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/174,365A
; FILING DATE: December 28, 1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 57:

Query Match 0.0%; Score 0; DB 1; Length 1;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 C 1

RESULT 4
US-08-174-365A-57
; Sequence 57, Application US/08174365A
; Patent No. 5478809
; GENERAL INFORMATION:
; APPLICANT: Seilichi TANIDA et al.
; TITLE OF INVENTION: 4-THIAHEPTANOIC ACID DERIVATIVES
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/174,365A
; FILING DATE: December 28, 1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 57:
```


SEQUENCE CHARACTERISTICS:
LENGTH: 1 amino acid
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: modified site
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: /note = "xaa is modified amino acid as
OTHER INFORMATION: described in specification"
US-08-174-365A-57

Query Match 0.0%; Score 0; DB 1; Length 1;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 X 1
DB 1 X 1

RESULT 5
US-07-789-913-23
; Sequence 23, Application US/07789913
; Patent No. 5559095

GENERAL INFORMATION:
APPLICANT: Miljanich, George P.
APPLICANT: Bowersox, Stephen S.
APPLICANT: Fox, James A.
APPLICANT: Valentino, Karen L.
APPLICANT: Bitner, Robert S.
APPLICANT: Yamashiro, Donald H.
TITLE OF INVENTION: Delayed Treatment Method of Reducing
TITLE OF INVENTION: Ischemia-Related Neuronal Damage
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Peter Dehlinger
STREET: 350 Cambridge Avenue, Suite 300
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/789,913
FILING DATE: 19911112
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/561,766
FILING DATE: 02-AUG-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/440,094
FILING DATE: 22-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Stratford, Carol A.
REGISTRATION NUMBER: 34,444
REFERENCE/DOCKET NUMBER: 5865-0005.30
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:
LENGTH: 1 amino acids
TYPE: AMINO ACID
TOPOLOGY: both
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO

ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: peptide fragment used in the claims
US-07-789-913-23

Query Match 0.0%; Score 0; DB 1; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 C 1

RESULT 6

US-07-789-913-25
; Sequence 25, Application US/07789913
; Patent No. 5559095

GENERAL INFORMATION:
APPLICANT: Miljanich, George P.
APPLICANT: Bowersox, Stephen S.
APPLICANT: Fox, James A.
APPLICANT: Valentino, Karen L.
APPLICANT: Bitner, Robert S.
APPLICANT: Yamashiro, Donald H.
TITLE OF INVENTION: Delayed Treatment Method of Reducing
TITLE OF INVENTION: Ischemia-Related Neuronal Damage
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Peter Dehlinger
STREET: 350 Cambridge Avenue, Suite 300
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/789,913
FILING DATE: 19911112
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/561,766
FILING DATE: 02-AUG-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/440,094
FILING DATE: 22-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Stratford, Carol A.
REGISTRATION NUMBER: 34,444
REFERENCE/DOCKET NUMBER: 5865-0005.30
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:
LENGTH: 1 amino acids
TYPE: AMINO ACID
TOPOLOGY: both
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: peptide fragment used in the claims

US-07-789-913-25

Query Match 0.0%; Score 0; DB 1; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1

Db 1 R 1

RESULT 7

US-08-049-794-23
; Sequence 23, Application US/08049794
; Patent No. 5587454
; GENERAL INFORMATION:

; APPLICANT: JUSTICE, ALAN
; APPLICANT: SINGH, TEJINDER
; APPLICANT: GOHIL, KISHOR C
; APPLICANT: VALENTINO, KAREN L
; APPLICANT: MILJANICH, GEORGE P
; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND
; TITLE OF INVENTION: ENHANCING OPIATE ANALGESIA
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Law Offices of Peter Dehlinger
; STREET: 350 Cambridge Avenue, Suite 300
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA

; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/049,794
; FILING DATE: 19930415
; CLASSIFICATION: 514
; PRIOR APPLICATION NUMBER: US 07/814,759
; FILING DATE: 30-DEC-1991
; ATTORNEY/AGENT INFORMATION:

; NAME: Stratford, Carol A.
; REGISTRATION NUMBER: 34,444
; REFERENCE/DOCKET NUMBER: 5865-0009.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960

; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear

; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT, PAGE
; INDIVIDUAL ISOLATE: 32
US-08-049-794-23

Query Match

Best Local Similarity 0.0%; Score 0; DB 1; Length 1;

Matches 0; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

Qy 1 X 1

Db 1 C 1

RESULT 8

US-08-049-794-25
; Sequence 25, Application US/08049794
; Patent No. 5587454
; GENERAL INFORMATION:

; APPLICANT: JUSTICE, ALAN
; APPLICANT: SINGH, TEJINDER
; APPLICANT: GOHIL, KISHOR C
; APPLICANT: VALENTINO, KAREN L

; APPLICANT: MILJANICH, GEORGE P
; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND
; TITLE OF INVENTION: ENHANCING OPIATE ANALGESIA
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Law Offices of Peter Dehlinger
; STREET: 350 Cambridge Avenue, Suite 300
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA

; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/049,794
; FILING DATE: 19930415
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/814,759
; FILING DATE: 30-DEC-1991

; ATTORNEY/AGENT INFORMATION:
; NAME: Stratford, Carol A.
; REGISTRATION NUMBER: 34,444
; REFERENCE/DOCKET NUMBER: 5865-0009.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960

; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear

; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT, PAGE
; INDIVIDUAL ISOLATE: 32
US-08-049-794-25

Query Match

Best Local Similarity 0.0%; Score 0; DB 1; Length 1;

Matches 0; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

Qy 1 X 1

Db 1 R 1

RESULT 9

US-08-433-037-12
; Sequence 12, Application US/08433037
; Patent No. 5707828
; GENERAL INFORMATION:

; APPLICANT: Sreekrishna, Kotikanyadan
; APPLICANT: Barr, Kathryn A.
; APPLICANT: Brierley, Russell A.
; APPLICANT: Thill, Gregory P.
; APPLICANT: Tschopp, Juerg F.

; TITLE OF INVENTION: EXPRESSION OF HUMAN SERUM ALBUMIN IN
; TITLE OF INVENTION: PICHIA PASTORIS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 11530-0299

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433.037
FILING DATE: 03-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 91082
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
US-08-433-037-12

Query Match 0.0%; Score 0; DB 1; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 L 1

RESULT 10
US-08-448-606-4
Sequence 4, Application US/08448606
Patent No. 5721114
GENERAL INFORMATION:
APPLICANT: Abrahams n, Lars
APPLICANT: Holmgren, Erik
APPLICANT: Kalder n, Christina
APPLICANT: Lake, Mats
APPLICANT: Mikaelsson, sa
APPLICANT: Sejlitz, Torsten
TITLE OF INVENTION: Expression System For Producing
TITLE OF INVENTION: Apolipoprotein AI-M
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pollock, Vande Sande & Priddy
STREET: 1990 M Street, N.W., Suite 800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/448,606
FILING DATE: 25-AUG-1995
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/SE93/01061
FILING DATE: 09-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 9203753-0
FILING DATE: 11-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Amernick, Burton A.
REGISTRATION NUMBER: 24,852
REFERENCE/DOCKET NUMBER: 0151/00121
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)331-7111
TELEFAX: (202) 293-6229
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1 amino acid
TYPE: amino acid
MOLECULE TYPE: peptide
US-08-448-606-4

Query Match 0.0%; Score 0; DB 1; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 Q 1

RESULT 11
US-07-869-933-16
Sequence 16, Application US/07869933
Patent No. 5770396
GENERAL INFORMATION:
APPLICANT: KINET, Jean-Pierre
TITLE OF INVENTION: ISOLATION, CHARACTERIZATION, AND USE OF
TITLE OF INVENTION: THE HUMAN B SUBUNIT OF THE HIGH AFFINITY RECEPTOR FOR
TITLE OF INVENTION: IMMUNOGLOBULIN
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/869,933
FILING DATE: 19920416
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 40399/154 NIHD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149

INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1 amino acids
TYPE: AMINO ACID
MOLECULE TYPE: protein
US-07-869-933-16

Query Match 0.0%; Score 0; DB 1; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 M 1

RESULT 12
US-08-293-150A-24
Sequence 24, Application US/08293150A

Patent No. 5792629
GENERAL INFORMATION:
APPLICANT: MORISHITA, Hideaki
APPLICANT: KANAMORI, Toshinori
APPLICANT: NOBUHARA, Masahiro
TITLE OF INVENTION: POLYPEPTIDE, DNA FRAGMENT ENCODING THE
TITLE OF INVENTION: SAME AND PROCESS FOR PRODUCING THE SAME, AND ENZYME
TITLE OF INVENTION: INHIBITION PROCESS, DRUG COMPOSITION AND METHODS OF
TITLE OF INVENTION: TREATING USING THE SAME
NUMBER OF SEQUENCES: 110
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/293,150A
FILING DATE: 19-AUG-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-306745
FILING DATE: 13-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 029650-049
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-293-150A-24

Query Match 0.0%; Score 0; DB 1; Length 1;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 C 1

RESULT 13
US-08-293-150A-40
Sequence 40, Application US/08293150A
Patent No. 5792629
GENERAL INFORMATION:
APPLICANT: MORISHITA, Hideaki
APPLICANT: KANAMORI, Toshinori
APPLICANT: NOBUHARA, Masahiro
TITLE OF INVENTION: POLYPEPTIDE, DNA FRAGMENT ENCODING THE
TITLE OF INVENTION: SAME AND PROCESS FOR PRODUCING THE SAME, AND ENZYME
TITLE OF INVENTION: INHIBITION PROCESS, DRUG COMPOSITION AND METHODS OF
TITLE OF INVENTION: TREATING USING THE SAME
NUMBER OF SEQUENCES: 110
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria

STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/293,150A
FILING DATE: 19-AUG-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/791,213
FILING DATE: 13-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-306745
FILING DATE: 13-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 029650-049
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-293-150A-40

Query Match 0.0%; Score 0; DB 1; Length 1;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 C 1

RESULT 14
US-08-496-847-23
Sequence 23, Application US/08496847
Patent No. 5795864
GENERAL INFORMATION:
APPLICANT: Amstutz, Gary A.
APPLICANT: Bowersox, Stephen S.
APPLICANT: Gohil, Kishorchandra
APPLICANT: Adriaenssens, Peter I.
APPLICANT: Kristipati, Ramasharma
TITLE OF INVENTION: METHODS AND
TITLE OF INVENTION: FORMULATIONS FOR PREVENTING PROGRESSION OF NEUROPATHIC PAIN
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94306-1546
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/496,847
FILING DATE: 27-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Stratford, Carol A

REGISTRATION NUMBER: 34,444
REFERENCE/DOCKET NUMBER: 5865-0009.31
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-0880
TELEFAX: 650-324-0960
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1 amino acid
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT
US-08-496-847-23

Query Match 0.0%; Score 0; DB 1; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 C 1

RESULT 15
US-08-496-847-25
Sequence 25, Application US/08496847
Patent No. 5795864
GENERAL INFORMATION:
APPLICANT: Amstutz, Gary A.
APPLICANT: Bowersox, Stephen S.
APPLICANT: Gohil, Kishorchandra
APPLICANT: Adriaenssens, Peter I.
APPLICANT: Kristipati, Ramasharma
TITLE OF INVENTION: FORMULATIONS FOR PREVENTING PROGRESSION OF NEUROPATHIC PAIN
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94306-1546
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Fastseq for Windows Version 2.0
CURRENT APPLICATION DATA:
FILING DATE: 27-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Stratford, Carol A.
REGISTRATION NUMBER: 34,444
REFERENCE/DOCKET NUMBER: 5865-0009.31
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-0880
TELEFAX: 650-324-0960
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1 amino acid
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT
US-08-496-847-25

Query Match 0.0%; Score 0; DB 1; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 C 1

Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 R 1

RESULT 16
US-08-742-774-23
Sequence 23, Application US/08742774
Patent No. 5824645
GENERAL INFORMATION:
APPLICANT: JUSTICE, ALAN
APPLICANT: SINGH, TEJINDER
APPLICANT: GOHIL, KISHOR C
APPLICANT: VALENTINO, KAREN L
APPLICANT: MILJANICH, GEORGE P
TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND
TITLE OF INVENTION: ENHANCING OPIATE ANALGESIA
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Peter Dehlinger
STREET: 350 Cambridge Avenue, Suite 300
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742,774
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/675,354
FILING DATE: 03-JUL-1996
APPLICATION NUMBER: US/08/049,794
FILING DATE: 1993-APR-15
APPLICATION NUMBER: US/07/814,759
FILING DATE: 30-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Stratford, Carol A.
REGISTRATION NUMBER: 34,444
REFERENCE/DOCKET NUMBER: 5865-0009.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1 amino acid
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT, PAGE
INDIVIDUAL ISOLATE: 32
US-08-742-774-23

Query Match 0.0%; Score 0; DB 2; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 C 1

RESULT 17

US-08-742-774-25
 ; Sequence 25, Application US/08742774
 ; Patent No. 5824645
 ; GENERAL INFORMATION:
 ; APPLICANT: JUSTICE, ALAN
 ; APPLICANT: SINGH, TEJINDER
 ; APPLICANT: GOHIL, KISHOR C
 ; APPLICANT: VALENTINO, KAREN L
 ; APPLICANT: MILJANICH, GEORGE P
 ; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND
 ; TITLE OF INVENTION: ENHANCING OPIATE ANALGESIA
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Law Offices of Peter Dehlinger
 ; STREET: 350 Cambridge Avenue, Suite 300
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94306
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/742,774
 ; FILING DATE: 30-DEC-1991
 ; NAME: Stratford, Carol A.
 ; REGISTRATION NUMBER: 34,444
 ; REFERENCE/DOCKET NUMBER: 5865-0009.30
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 324-0880
 ; TELEFAX: (415) 324-0960
 ; INFORMATION FOR SEQ ID NO: 25:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ORIGINAL SOURCE:
 ; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT, PAGE
 ; INDIVIDUAL ISOLATE: 32
 ; US-08-742-774-25

Query Match 0.0%; Score 0; DB 2; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

QY 1 X 1
 Db 1 R 1

RESULT 18
 US-08-675-354-23
 ; Sequence 25, Application US/08675354
 ; Patent No. 5859186
 ; GENERAL INFORMATION:
 ; APPLICANT: JUSTICE, ALAN
 ; APPLICANT: SINGH, TEJINDER
 ; APPLICANT: GOHIL, KISHOR C
 ; APPLICANT: VALENTINO, KAREN L
 ; APPLICANT: MILJANICH, GEORGE P
 ; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND
 ; TITLE OF INVENTION: ENHANCING OPIATE ANALGESIA
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Law Offices of Peter Dehlinger
 ; STREET: 350 Cambridge Avenue, Suite 300
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94306
 ; COMPUTER READABLE FORM:

US-08-675-354-25
 ; Sequence 25, Application US/08675354
 ; Patent No. 5859186
 ; GENERAL INFORMATION:
 ; APPLICANT: JUSTICE, ALAN
 ; APPLICANT: SINGH, TEJINDER
 ; APPLICANT: GOHIL, KISHOR C
 ; APPLICANT: VALENTINO, KAREN L
 ; APPLICANT: MILJANICH, GEORGE P
 ; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND
 ; TITLE OF INVENTION: ENHANCING OPIATE ANALGESIA
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Law Offices of Peter Dehlinger
 ; STREET: 350 Cambridge Avenue, Suite 300
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94306
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/675,354
 ; FILING DATE: 03-JUL-1996
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/049,794
 ; FILING DATE: 1993-APR-15
 ; APPLICATION NUMBER: US 07/814,759
 ; FILING DATE: 30-DEC-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Stratford, Carol A.
 ; REGISTRATION NUMBER: 34,444
 ; REFERENCE/DOCKET NUMBER: 5865-0009.30
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 324-0880
 ; TELEFAX: (415) 324-0960
 ; INFORMATION FOR SEQ ID NO: 23:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ORIGINAL SOURCE:
 ; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT, PAGE
 ; INDIVIDUAL ISOLATE: 32
 ; US-08-675-354-23

Query Match 0.0%; Score 0; DB 2; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

QY 1 X 1
 Db 1 C 1

RESULT 19
 US-08-675-354-25
 ; Sequence 25, Application US/08675354
 ; Patent No. 5859186
 ; GENERAL INFORMATION:
 ; APPLICANT: JUSTICE, ALAN
 ; APPLICANT: SINGH, TEJINDER
 ; APPLICANT: GOHIL, KISHOR C
 ; APPLICANT: VALENTINO, KAREN L
 ; APPLICANT: MILJANICH, GEORGE P
 ; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND
 ; TITLE OF INVENTION: ENHANCING OPIATE ANALGESIA
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Law Offices of Peter Dehlinger
 ; STREET: 350 Cambridge Avenue, Suite 300
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94306
 ; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/675,354
;; FILING DATE: 03-JUL-1996
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/049,794
;; FILING DATE: 1993-APR-15
;; APPLICATION NUMBER: US 07/814,759
;; FILING DATE: 30-DEC-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Stratford, Carol A.
;; REGISTRATION NUMBER: 34,444
;; REFERENCE/DOCKET NUMBER: 5865-0009.30
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 324-0880
;; TELEFAX: (415) 324-0960
;; INFORMATION FOR SEQ ID NO: 25:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHETICAL: NO
;; ORIGINAL SOURCE:
;; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT, PAGE
;; INDIVIDUAL ISOLATE: 32
US-08-675-354-25

Query Match 0.0%; Score 0; DB 2; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1

Db 1 R 1

RESULT 20
US-08-097-554A-12
;; Sequence 12, Application US/08097554A
;; Patent No. 5869312
;; GENERAL INFORMATION:
;; APPLICANT: Cochran Ph.D., Mark D
;; APPLICANT: Junker M.S., David E
;; TITLE OF INVENTION: Recombinant Swinepox Virus
;; NUMBER OF SEQUENCES: 112
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: John P. White
;; STREET: 30 Rockefeller Plaza
;; CITY: New York
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10112
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/097,554A
;; FILING DATE: July 22, 1993
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: White, John P
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 977-9550
;; TELEFAX: (212) 664-0525
;; TELEX: 422523
;; INFORMATION FOR SEQ ID NO: 12:

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-097-554A-12
Query Match 0.0%; Score 0; DB 2; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1

Db 1 V 1

RESULT 21
US-08-965-918-23
;; Sequence 23, Application US/08965918
;; Patent No. 5891849
;; GENERAL INFORMATION:
;; APPLICANT: Amstutz, Gary A.
;; APPLICANT: Bowersox, Stephen S.
;; APPLICANT: Gohil, Kishorchandra
;; APPLICANT: Adriaenssens, Peter I.
;; APPLICANT: Kristipati, Ramasharma
;; TITLE OF INVENTION: METHODS AND FORMULATIONS FOR PREVENTING
;; TITLE OF INVENTION: PROGRESSION OF NEUROPATHIC PAIN
;; NUMBER OF SEQUENCES: 36
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Dehlinger & Associates
;; STREET: 350 Cambridge Avenue, Suite 250
;; CITY: Palo Alto
;; STATE: CA
;; COUNTRY: US
;; ZIP: 94306-1546
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/965,918
;; FILING DATE: 07-NOV-1997
;; CLASSIFICATION: 514
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Mohr, Judy M.
;; REGISTRATION NUMBER: 38,563
;; REFERENCE/DOCKET NUMBER: 5865-0009.34
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 650-324-0880
;; TELEFAX: 650-324-0960
;; INFORMATION FOR SEQ ID NO: 23:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHETICAL: NO
;; ORIGINAL SOURCE:
;; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT
US-08-965-918-23

Query Match

Best Local Similarity 0.0%; Score 0; DB 2; Length 1;

Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1

Db 1 C 1

RESULT 22

US-08-965-918-25
; Sequence 25, Application US/08965918
; Patent No. 5891849
; GENERAL INFORMATION:
; APPLICANT: Amstutz, Gary A.
; APPLICANT: Bowersox, Stephen S.
; APPLICANT: Gohil, Kishorchandra
; APPLICANT: Adriaenssens, Peter I.
; APPLICANT: Ktistipati, Ramasharma
; TITLE OF INVENTION: METHODS AND FORMULATIONS FOR PREVENTING
; TITLE OF INVENTION: PROGRESSION OF NEUROPATHIC PAIN
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94306-1546
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/965,918
; FILING DATE: 07-NOV-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohr, Judy M.
; REGISTRATION NUMBER: 38,563
; REFERENCE/DOCKET NUMBER: 5865-0009.34
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0880
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT
US-08-965-918-25
Query Match 0.0%; Score 0; DB 2; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 X 1
Db 1 R 1
RESULT 23
US-09-138-439-23
; Sequence 23, Application US/09138439
; Patent No. 5994305
; GENERAL INFORMATION:
; APPLICANT: JUSTICE, ALAN
; APPLICANT: SINGH, TEJINDER
; APPLICANT: GOHIL, KISHOR C
; APPLICANT: VALENTINO, KAREN L
; APPLICANT: MILJANICH, GEORGE P
; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND
; TITLE OF INVENTION: ENHANCING OPIATE ANALGESIA
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Peter Dehlinger
; STREET: 350 Cambridge Avenue, Suite 300
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/138,439
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/049,794
; FILING DATE: 1993-04-15
; APPLICATION NUMBER: US 07/814,759
; FILING DATE: 30-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Stratford, Carol A.
; REGISTRATION NUMBER: 34,444
; REFERENCE/DOCKET NUMBER: 5865-0009.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT, PAGE

US-09-138-439-23
Query Match 0.0%; Score 0; DB 2; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 X 1
Db 1 C 1
RESULT 24
US-09-138-439-25
; Sequence 25, Application US/09138439
; Patent No. 5994305
; GENERAL INFORMATION:
; APPLICANT: JUSTICE, ALAN
; APPLICANT: SINGH, TEJINDER
; APPLICANT: GOHIL, KISHOR C
; APPLICANT: VALENTINO, KAREN L
; APPLICANT: MILJANICH, GEORGE P
; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND
; TITLE OF INVENTION: ENHANCING OPIATE ANALGESIA
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Peter Dehlinger
; STREET: 350 Cambridge Avenue, Suite 300
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/138,439
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/049,794
; FILING DATE: 1993-04-15
; APPLICATION NUMBER: US 07/814,759
; FILING DATE: 30-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Stratford, Carol A.
; REGISTRATION NUMBER: 34,444
; REFERENCE/DOCKET NUMBER: 5865-0009.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT, PAGE

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/049,794
;; FILING DATE: 1993-04-15
;; APPLICATION NUMBER: US 07/814,759
;; FILING DATE: 30-DEC-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Stratford, Carol A.
;; REGISTRATION NUMBER: 34,444
;; REFERENCE/DOCKET NUMBER: 5865-0009.30
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 324-0880
;; TELEFAX: (415) 324-0960
;; INFORMATION FOR SEQ ID NO: 25:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHETICAL: NO
;; ORIGINAL SOURCE:
;; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT, PAGE
;; INDIVIDUAL ISOLATE: 32
US-09-138-439-25

Query Match 0.0%; Score 0; DB 2; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 R 1

RESULT 25
US-08-480-640A-12
; Sequence 12, Application US/08480640A
; Patent No. 6033904
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,640A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-480-640A-12

Query Match 0.0%; Score 0; DB 3; Length 1;

Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 V 1

RESULT 26
US-08-613-400A-23
; Sequence 23, Application US/08613400A
; Patent No. 6054429
; GENERAL INFORMATION:
; APPLICANT: Bowersox, S. Scott
; APPLICANT: Gadbois, Theresa
; APPLICANT: Pettus, Mark, R.
; APPLICANT: Luther, Robert, R.
; TITLE OF INVENTION: IMPROVED EPIDURAL
; TITLE OF INVENTION: METHOD OF PRODUCING ANALGESIA
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94306-1546
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/613,400A
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stratford, Carol A.
; REGISTRATION NUMBER: 34,444
; REFERENCE/DOCKET NUMBER: 5865-0019
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0880
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT,
; INDIVIDUAL ISOLATE: PAGE 33, LINES 16-28
US-08-613-400A-23

Query Match 0.0%; Score 0; DB 3; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 C 1

RESULT 27
US-08-613-400A-25
; Sequence 25, Application US/08613400A
; Patent No. 6054429
; GENERAL INFORMATION:
; APPLICANT: Bowersox, S. Scott

APPLICANT: Gadbois, Theresa
APPLICANT: Pettus, Mark, R.
APPLICANT: Luther, Robert, R.
TITLE OF INVENTION: IMPROVED EPIDURAL
TITLE OF INVENTION: METHOD OF PRODUCING ANALGESIA
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94306-1546
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/613,400A
FILING DATE: 08-MAR-1996
CLASSIFICATION: 514
PRIOR APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Stratford, Carol A
REGISTRATION NUMBER: 34,444
REFERENCE/DOCKET NUMBER: 5865-0019
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-0880
TELEFAX: 650-324-0960
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT,
INDIVIDUAL ISOLATE: PAGE 33, LINES 16-28
US-08-613-400A-25

Query Match 0.0%; Score 0; DB 3; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 R 1

RESULT 28
US-08-801-092-10
Sequence 10, Application US/08801092
Patent No. 6074850
GENERAL INFORMATION:
APPLICANT: Antelman, Douglas
APPLICANT: Gregory, Richard J.
APPLICANT: Wils, Kenneth N.
TITLE OF INVENTION: Tissue Specific Expression of
RETINOBLASTOMA PROTEIN
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/801,092
FILING DATE: 14-FEB-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/751,517
FILING DATE: 15-NOV-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 016930-001020
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 703-576-0300
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-801-092-10

Query Match 0.0%; Score 0; DB 3; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 R 1

RESULT 29
US-08-801-092-17
Sequence 17, Application US/08801092
Patent No. 6074850
GENERAL INFORMATION:
APPLICANT: Antelman, Douglas
APPLICANT: Gregory, Richard J.
APPLICANT: Wils, Kenneth N.
TITLE OF INVENTION: Tissue Specific Expression of
RETINOBLASTOMA PROTEIN
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/801,092
FILING DATE: 14-FEB-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/751,517
FILING DATE: 15-NOV-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 016930-001020
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 703-576-0300

; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-801-092-17

Query Match 0.0%; Score 0; DB 3; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 L 1

RESULT 30

US-08-801-092-24
; Sequence 24, Application US/08801092

; Patent No. 6074850
; GENERAL INFORMATION:
; APPLICANT: Antelman, Douglas
; APPLICANT: Gregory, Richard J.
; APPLICANT: Wils, Kenneth N.
; TITLE OF INVENTION: Tissue Specific Expression of
; TITLE OF INVENTION: Retinoblastoma Protein
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,092
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/751,517
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 016930-001020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 703-576-0300

; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-801-092-24

Query Match 0.0%; Score 0; DB 3; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 R 1

RESULT 31

US-08-801-092-31

; Sequence 31, Application US/08801092
; Patent No. 6074850
; GENERAL INFORMATION:
; APPLICANT: Antelman, Douglas
; APPLICANT: Gregory, Richard J.
; APPLICANT: Wils, Kenneth N.
; TITLE OF INVENTION: Tissue Specific Expression of
; TITLE OF INVENTION: Retinoblastoma Protein
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,092
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/751,517
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 016930-001020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 703-576-0300

; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-801-092-31

Query Match 0.0%; Score 0; DB 3; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 L 1

RESULT 32

US-08-801-092-38
; Sequence 38, Application US/08801092

; Patent No. 6074850
; GENERAL INFORMATION:
; APPLICANT: Antelman, Douglas
; APPLICANT: Gregory, Richard J.
; APPLICANT: Wils, Kenneth N.
; TITLE OF INVENTION: Tissue Specific Expression of
; TITLE OF INVENTION: Retinoblastoma Protein
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111

; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/801,092
FILING DATE: 14-FEB-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/751,517
FILING DATE: 15-NOV-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 016930-001020
TELEPHONE: 415-576-0200
TELEFAX: 703-576-0300
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 1 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-801-092-38

Query Match 0.0%; Score 0; DB 3; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 R 1

RESULT 33
US-08-801-092-45
Sequence 45, Application US/08801092
Patent No. 6074850
GENERAL INFORMATION:
APPLICANT: Antelman, Douglas
APPLICANT: Gregory, Richard J.
APPLICANT: Wils, Kenneth N.
TITLE OF INVENTION: Tissue Specific Expression of
TITLE OF INVENTION: Retinoblastoma Protein
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/801,092
FILING DATE: 14-FEB-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/751,517
FILING DATE: 15-NOV-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 016930-001020
TELEPHONE: 415-576-0200

TELEFAX: 703-576-0300
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 1 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-801-092-45

Query Match 0.0%; Score 0; DB 3; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 L 1

RESULT 34
US-09-298-017-23
Sequence 23, Application US/09298017
Patent No. 6087091
GENERAL INFORMATION:
APPLICANT: JUSTICE, ALAN
APPLICANT: SINGH, TEJINDER
APPLICANT: GOHIL, KISHOR C
APPLICANT: VALENTINO, KAREN L
APPLICANT: MILJANICH, GEORGE P
TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND
TITLE OF INVENTION: ENHANCING OPIATE ANALGESIA
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Peter Dehlinger
STREET: 350 Cambridge Avenue, Suite 300
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/298,017
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/049,794
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Stratford, Carol A.
REGISTRATION NUMBER: 34,444
REFERENCE/DOCKET NUMBER: 5865-0009.30
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE:
INDIVIDUAL ISOLATE: 32
US-09-298-017-23

Query Match 0.0%; Score 0; DB 3; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 X 1
Db 1 C 1

RESULT 35
US-09-298-017-25
; Sequence 25, Application US/09298017
; Patent No. 6087091
; GENERAL INFORMATION:
; APPLICANT: JUSTICE, ALAN
; APPLICANT: SINGH, TEJINDER
; APPLICANT: GOHIL, KISHOR C
; APPLICANT: VALENTINO, KAREN L
; APPLICANT: MILJANICH, GEORGE P
; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND
; TITLE OF INVENTION: ENHANCING OPIATE ANALGESIA
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Peter Dehlinger
; STREET: 350 Cambridge Avenue, Suite 300
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/298,017
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/049,794
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stratford, Carol A.
; REGISTRATION NUMBER: 34,444
; REFERENCE/DOCKET NUMBER: 5865-0009.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT, PAGE
; INDIVIDUAL ISOLATE: 32
US-09-298-017-25
Query Match 0.0%; Score 0; DB 3; Length 1;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 1 X 1
Db 1 R 1

RESULT 36
US-08-295-802-12
; Sequence 12, Application US/08295802
; Patent No. 6127163
; GENERAL INFORMATION:
; APPLICANT: Cochran Ph.D., Mark D
; APPLICANT: Junker M.S., David E
; TITLE OF INVENTION: Recombinant Swinepox Virus

US-09-998-861-14.ra1
; NUMBER OF SEQUENCES: 188
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/295,802
; FILING DATE: Herewith
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)977-9550
; TELEFAX: (212)664-0525
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-295-802-12
Query Match 0.0%; Score 0; DB 3; Length 1;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 1 X 1
Db 1 V 1

RESULT 37
US-09-392-979A-23
; Sequence 23, Application US/09392979A
; Patent No. 6136786
; GENERAL INFORMATION:
; APPLICANT: JUSTICE, ALAN
; APPLICANT: SINGH, TEJINDER
; APPLICANT: GOHIL, KISHOR C
; APPLICANT: VALENTINO, KAREN L
; APPLICANT: MILJANICH, GEORGE P
; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND
; TITLE OF INVENTION: ENHANCING OPIATE ANALGESIA
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Peter Dehlinger
; STREET: 350 Cambridge Avenue, Suite 300
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/392,979A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/049,794
; FILING DATE: 1993-04-15
; APPLICATION NUMBER: US 07/814,759

;; FILING DATE: 30-DEC-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Stratford, Carol A.
;; REGISTRATION NUMBER: 34,444
;; REFERENCE/DOCKET NUMBER: 5865-0009.30
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 324-0880
;; TELEFAX: (415) 324-0960
;; INFORMATION FOR SEQ ID NO: 23:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHETICAL: NO
;; ORIGINAL SOURCE:
;; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT, PAGE 32
;; INDIVIDUAL ISOLATE: 32
US-09-392-979A-23

Query Match 0.0%; Score 0; DB 3; Length 1;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 C 1

RESULT 38
US-09-392-979A-25
; Sequence 25, Application US/09392979A
; Patent No. 6136786
; GENERAL INFORMATION:
; APPLICANT: JUSTICE, ALAN
; APPLICANT: SINGH, TEJINDER
; APPLICANT: GOHIL, KISHOR C
; APPLICANT: VALENTINO, KAREN L
; APPLICANT: MILJANICH, GEORGE P
; TITLE OF INVENTION: METHODS OF PRODUCING ANALGESIA AND
; ENHANCING OPIATE ANALGESIA
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; STREET: Law Offices of Peter Dehlinger
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/09/392,979A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/049,794
; FILING DATE: 1993-04-15
; APPLICATION NUMBER: US 07/814,759
; FILING DATE: 30-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Stratford, Carol A.
; REGISTRATION NUMBER: 34,444
; REFERENCE/DOCKET NUMBER: 5865-0009.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1 amino acids

;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHETICAL: NO
;; ORIGINAL SOURCE:
;; INDIVIDUAL ISOLATE: CONOPEPTIDE GROUP 1 FRAGMENT, PAGE 32
;; INDIVIDUAL ISOLATE: 32
US-09-392-979A-25

Query Match 0.0%; Score 0; DB 3; Length 1;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 R 1

RESULT 39
US-09-103-663-16
; Sequence 16, Application US/09103663D
; Patent No. 6171803
; GENERAL INFORMATION:
; APPLICANT: Kinet et al.
; TITLE OF INVENTION: Isolation, characterization, and use of the human beta
; TITLE OF INVENTION: subunit of the high affinity receptor for
; TITLE OF INVENTION: immunoglobulin E.
; FILE REFERENCE: 50490
; CURRENT APPLICATION NUMBER: US/09/103,663D
; CURRENT FILING DATE: 1998-06-23
; EARLIER APPLICATION NUMBER: 07/869,933
; EARLIER FILING DATE: 1992-04-16
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-103-663-16

Query Match 0.0%; Score 0; DB 3; Length 1;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 M 1

RESULT 40
US-08-488-237A-12
; Sequence 12, Application US/08488237A
; Patent No. 6251403
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,237A
; FILING DATE: 07-JUN-1995


```

; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-488-237A-12

Query Match      0.0%; Score 0; DB 3; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 X 1
Db      1 V 1

```

Search completed: April 13, 2005, 17:18:52
Job time : 3.86149 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2005, 17:06:25 ; Search time 8.32432 Seconds
(Without alignments)
439.204 Million cell updates/sec

Title: US-09-998-861-14
Perfect score: 11
Sequence: 1 XXXXXXXXXX 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1421835 seqs, 332370683 residues

Total number of hits satisfying chosen parameters: 1421835

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	0	0.0	1	9	US-09-778-885-5
2	0	0.0	1	9	US-09-909-348-4
3	0	0.0	1	9	US-09-982-172-3
4	0	0.0	1	9	US-09-982-172-4
5	0	0.0	1	9	US-09-982-172-9
6	0	0.0	1	9	US-09-982-172-11
7	0	0.0	1	9	US-09-982-172-19
8	0	0.0	1	9	US-09-982-172-31
9	0	0.0	1	9	US-09-982-172-35
10	0	0.0	1	9	US-09-982-172-37
11	0	0.0	1	9	US-09-982-172-46
12	0	0.0	1	9	US-09-982-172-69
13	0	0.0	1	9	US-09-982-172-80

14	0	0.0	1	9	US-09-982-172-81	Sequence 81, Appl
15	0	0.0	1	9	US-09-982-172-83	Sequence 83, Appl
16	0	0.0	1	9	US-09-982-172-86	Sequence 86, Appl
17	0	0.0	1	9	US-09-982-172-93	Sequence 93, Appl
18	0	0.0	1	9	US-09-982-172-95	Sequence 95, Appl
19	0	0.0	1	9	US-09-982-172-106	Sequence 106, Appl
20	0	0.0	1	9	US-09-982-172-112	Sequence 112, Appl
21	0	0.0	1	9	US-09-982-172-120	Sequence 120, Appl
22	0	0.0	1	9	US-09-982-172-126	Sequence 126, Appl
23	0	0.0	1	9	US-09-982-172-148	Sequence 148, Appl
24	0	0.0	1	9	US-09-982-172-149	Sequence 149, Appl
25	0	0.0	1	9	US-09-982-172-155	Sequence 155, Appl
26	0	0.0	1	9	US-09-982-172-160	Sequence 160, Appl
27	0	0.0	1	9	US-09-982-172-172	Sequence 172, Appl
28	0	0.0	1	9	US-09-982-172-173	Sequence 173, Appl
29	0	0.0	1	9	US-09-982-172-175	Sequence 175, Appl
30	0	0.0	1	9	US-09-982-172-189	Sequence 189, Appl
31	0	0.0	1	9	US-09-982-172-190	Sequence 190, Appl
32	0	0.0	1	9	US-09-982-172-191	Sequence 191, Appl
33	0	0.0	1	9	US-09-982-172-195	Sequence 195, Appl
34	0	0.0	1	9	US-09-982-172-200	Sequence 200, Appl
35	0	0.0	1	9	US-09-982-172-211	Sequence 211, Appl
36	0	0.0	1	10	US-09-809-391-395	Sequence 395, Appl
37	0	0.0	1	10	US-09-809-391-611	Sequence 611, Appl
38	0	0.0	1	10	US-09-882-171-395	Sequence 395, Appl
39	0	0.0	1	10	US-09-882-171-611	Sequence 611, Appl
40	0	0.0	1	11	US-09-833-245-184	Sequence 184, Appl
41	0	0.0	1	11	US-09-833-245-186	Sequence 186, Appl
42	0	0.0	1	11	US-09-833-245-325	Sequence 325, Appl
43	0	0.0	1	11	US-09-833-245-500	Sequence 500, Appl
44	0	0.0	1	11	US-09-833-245-744	Sequence 744, Appl
45	0	0.0	1	11	US-09-833-245-1045	Sequence 1045, Ap

ALIGNMENTS

RESULT 1
US-09-778-885-5
; Sequence 5, Application US/09778885
; Publication No. US20020039748A1
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Foster, Donald C.
; APPLICANT: Gao, Zeren
; TITLE OF INVENTION: KUNITZ DOMAIN POLYPEPTIDE AND MATERIALS
; TITLE OF INVENTION: AND METHODS FOR MAKING IT
; FILE REFERENCE: 98-22
; CURRENT APPLICATION NUMBER: US/09/778,885
; CURRENT FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: 09/320,095
; PRIOR FILING DATE: 1999-05-26
; PRIOR APPLICATION NUMBER: US 60/087,032
; PRIOR FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: polypeptide motif
; NAME/KEY: VARIANT
; LOCATION: (2)...(2)
; OTHER INFORMATION: Xaa is any residue except Asp, Cys, Gly, His, Met.
; OTHER INFORMATION: Pro, Trp or Val
; NAME/KEY: VARIANT
; LOCATION: (3)...(3)
; OTHER INFORMATION: Xaa is Leu, Glu, Met, Gln, Phe, Ser, Thr, Ala or Pro
; NAME/KEY: VARIANT
; LOCATION: (4)...(4)
; OTHER INFORMATION: Xaa is any residue except Arg, Cys, Met, Phe, Trp.
; OTHER INFORMATION: Tyr or Val


```

; NAME/KEY: VARIANT
; LOCATION: (5)...(5)
; OTHER INFORMATION: Xaa is any residue except Asn, Cys, Gln, Gly, Phe,
; OTHER INFORMATION: Ser, Thr or Trp
; NAME/KEY: VARIANT
; LOCATION: (6)...(6)
; OTHER INFORMATION: Xaa is Arg, Glu, Asn, Ala, Val, Asp, Lys, Ser, Tyr
; OTHER INFORMATION: or Met
; NAME/KEY: VARIANT
; LOCATION: (7)...(7)
; OTHER INFORMATION: Xaa is any residue except Asn, Cys, Gly, His, Leu,
; OTHER INFORMATION: Met, Phe or Trp
; NAME/KEY: VARIANT
; LOCATION: (8)...(8)
; OTHER INFORMATION: Xaa is Gly or Glu
; NAME/KEY: VARIANT
; LOCATION: (9)...(9)
; OTHER INFORMATION: Xaa is Pro, Arg, Leu, Val, Ser, Asp, Ile, Asn or Thr
; NAME/KEY: VARIANT
; LOCATION: (11)...(11)
; OTHER INFORMATION: Xaa is any residue except Ala, Cys, Glu, His, Ile,
; OTHER INFORMATION: Pro, Trp and Val
; NAME/KEY: VARIANT
; LOCATION: (12)...(12)
; OTHER INFORMATION: Xaa is Arg, Lys, Ala, Asp, Gln, Phe, Gly, Glu, Thr
; OTHER INFORMATION: and Ser
; NAME/KEY: VARIANT
; LOCATION: (13)...(13)
; OTHER INFORMATION: Xaa is any residue except Asp, Cys, Glu, Pro, Thr
; OTHER INFORMATION: or Trp
; NAME/KEY: VARIANT
; LOCATION: (14)...(14)
; OTHER INFORMATION: Xaa is any residue except Arg, Asn, Cys, Gly, His,
; OTHER INFORMATION: Ser, Trp or Tyr
; NAME/KEY: VARIANT
; LOCATION: (15)...(15)
; OTHER INFORMATION: Xaa is any residue except Ala, Asp, Cys, Gly, His,
; OTHER INFORMATION: Met, Trp or Tyr
; NAME/KEY: VARIANT
; LOCATION: (16)...(16)
; OTHER INFORMATION: Xaa is Ser, Ala, Arg, Val, Gln, Lys, Leu, Gly or Ile
; NAME/KEY: VARIANT
; LOCATION: (17)...(17)
; OTHER INFORMATION: Xaa is Phe, Tyr, Ile, Trp or Leu
; NAME/KEY: VARIANT
; LOCATION: (18)...(18)
; OTHER INFORMATION: Xaa is Tyr, His, Phe, Trp, Asn or Ala
; NAME/KEY: VARIANT
; LOCATION: (19)...(19)
; OTHER INFORMATION: Xaa is Tyr or Phe
; NAME/KEY: VARIANT
; LOCATION: (20)...(20)
; OTHER INFORMATION: Xaa is Lys, Asn, Ser or Asp
; NAME/KEY: VARIANT
; LOCATION: (21)...(21)
; OTHER INFORMATION: Xaa is any residue except Asp, Cys, Glu, His or Tyr
; NAME/KEY: VARIANT
; LOCATION: (22)...(22)
; OTHER INFORMATION: Xaa is any residue except Cys, Met, Pro or Trp
; NAME/KEY: VARIANT
; LOCATION: (23)...(23)
; OTHER INFORMATION: Xaa is Ala, Lys, Ser, Leu, Thr, Ile, Gln, Glu, Tyr or Val
; NAME/KEY: VARIANT
; LOCATION: (24)...(24)
; OTHER INFORMATION: Xaa is Lys, Gln, Asn, His, Gly, Arg or Met
; NAME/KEY: VARIANT
; LOCATION: (25)...(25)
; OTHER INFORMATION: Xaa is any residue except Asn, Asp, Cys, His, Ile, Pro, Trp, Tyr
; OTHER INFORMATION: Xaa is any residue except Asn, Asp, Cys, His, Ile, Pro, Trp, Tyr
US-09-778-885-5

```

Query Match 0.0%; Score 0; DB 9; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
 Db 1 V 1

RESULT 2
 US-09-909-348-4
 ; Sequence 4, Application US/09909348
 ; Patent No. US20020042373A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Carney, Darrell H.
 ; APPLICANT: Crowther, Roger S.
 ; APPLICANT: Stierberg, Janet
 ; APPLICANT: Bergmann, John
 ; TITLE OF INVENTION: Stimulation Of Cartilage Growth With Agonists
 ; FILE REFERENCE: 3033.1003-001
 ; CURRENT APPLICATION NUMBER: US/09/909,348
 ; CURRENT FILING DATE: 2001-07-19
 ; PRIOR APPLICATION NUMBER: US 60/219,800
 ; PRIOR FILING DATE: 2000-07-20
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 14
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Peptide fragment of Thrombin
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)...(14)
 ; OTHER INFORMATION: Xaa at position six is Glu or Gln
 ; OTHER INFORMATION: Xaa at position thirteen is Phe, Met, Leu, His or
 US-09-909-348-4

Query Match 0.0%; Score 0; DB 9; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
 Db 1 V 1

RESULT 3
 US-09-982-172-3
 ; Sequence 3, Application US/09982172
 ; Patent No. US20020137119A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Emil Israel Katz
 ; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODI
 ; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERAT
 ; TITLE OF INVENTION: UTILIZING EACH
 ; FILE REFERENCE: 01/22283
 ; CURRENT APPLICATION NUMBER: US/09/982,172
 ; CURRENT FILING DATE: 2001-10-19
 ; NUMBER OF SEQ ID NOS: 253
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 3
 ; LENGTH: 1
 ; TYPE: PRT
 ; ORGANISM: Artificial sequence
 ; FEATURE:
 ; OTHER INFORMATION: Computer generated synthetic peptide
 US-09-982-172-3

Query Match 0.0%; Score 0; DB 9; Length 1;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1

Db 1 K 1

RESULT 4

US-09-982-172-4

; Sequence 4, Application US/09982172

; Patent No. US20020137119A1

; GENERAL INFORMATION:

; APPLICANT: Emil Israel Katz

; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES

; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING

; FILE REFERENCE: 01/22283

; CURRENT APPLICATION NUMBER: US/09/982,172

; CURRENT FILING DATE: 2001-10-19

; NUMBER OF SEQ ID NOS: 253

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 4

; LENGTH: 1

; TYPE: PRT

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: Computer generated synthetic peptide

US-09-982-172-4

Query Match

Best Local Similarity 0.0%; Score 0; DB 9; Length 1;

Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1

Db 1 K 1

RESULT 5

US-09-982-172-9

; Sequence 9, Application US/09982172

; Patent No. US20020137119A1

; GENERAL INFORMATION:

; APPLICANT: Emil Israel Katz

; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES

; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING

; FILE REFERENCE: 01/22283

; CURRENT APPLICATION NUMBER: US/09/982,172

; CURRENT FILING DATE: 2001-10-19

; NUMBER OF SEQ ID NOS: 253

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 9

; LENGTH: 1

; TYPE: PRT

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: Computer generated synthetic peptide

US-09-982-172-9

Query Match

Best Local Similarity 0.0%; Score 0; DB 9; Length 1;

Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1

Db 1 K 1

RESULT 6

US-09-982-172-11

; Sequence 11, Application US/09982172

; Patent No. US20020137119A1

; GENERAL INFORMATION:

; APPLICANT: Emil Israel Katz

; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES

; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING

; TITLE OF INVENTION: UTILIZING EACH

; FILE REFERENCE: 01/22283

; CURRENT APPLICATION NUMBER: US/09/982,172

; CURRENT FILING DATE: 2001-10-19

; NUMBER OF SEQ ID NOS: 253

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 11

; LENGTH: 1

; TYPE: PRT

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: Computer generated synthetic peptide

US-09-982-172-11

Query Match

Best Local Similarity 0.0%; Score 0; DB 9; Length 1;

Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1

Db 1 K 1

RESULT 7

US-09-982-172-19

; Sequence 19, Application US/09982172

; Patent No. US20020137119A1

; GENERAL INFORMATION:

; APPLICANT: Emil Israel Katz

; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES

; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING

; FILE REFERENCE: 01/22283

; CURRENT APPLICATION NUMBER: US/09/982,172

; CURRENT FILING DATE: 2001-10-19

; NUMBER OF SEQ ID NOS: 253

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 19

; LENGTH: 1

; TYPE: PRT

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: Computer generated synthetic peptide

US-09-982-172-19

Query Match

Best Local Similarity 0.0%; Score 0; DB 9; Length 1;

Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1

Db 1 K 1

RESULT 8

US-09-982-172-31

; Sequence 31, Application US/09982172

; Patent No. US20020137119A1

; GENERAL INFORMATION:

; APPLICANT: Emil Israel Katz

; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES

; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING

; FILE REFERENCE: 01/22283

; CURRENT APPLICATION NUMBER: US/09/982,172

; CURRENT FILING DATE: 2001-10-19

; NUMBER OF SEQ ID NOS: 253

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 31

; LENGTH: 1

; TYPE: PRT

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-31

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 K 1

RESULT 9

US-09-982-172-35
; Sequence 35, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:

; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35

; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide

US-09-982-172-35

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 R 1

RESULT 10

US-09-982-172-37
; Sequence 37, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:

; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37

; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide

US-09-982-172-37

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 K 1

RESULT 11

US-09-982-172-46
; Sequence 46, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46

; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide

US-09-982-172-46

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 K 1

RESULT 12

US-09-982-172-69
; Sequence 69, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:

; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69

; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide

US-09-982-172-69

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 K 1

RESULT 13

US-09-982-172-80
; Sequence 80, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:

; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283

; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 80
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-80

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 K 1

RESULT 14

US-09-982-172-81
; Sequence 81, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 81
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-81

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 R 1

RESULT 15

US-09-982-172-83
; Sequence 83, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 83
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-83

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 R 1

RESULT 16

US-09-982-172-86

; Sequence 86, Application US/09982172

; Patent No. US20020137119A1

; GENERAL INFORMATION:

; APPLICANT: Emil Israel Katz

; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES

; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING

; TITLE OF INVENTION: UTILIZING EACH

; FILE REFERENCE: 01/22283

; CURRENT APPLICATION NUMBER: US/09/982,172

; CURRENT FILING DATE: 2001-10-19

; NUMBER OF SEQ ID NOS: 253

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 86

; LENGTH: 1

; TYPE: PRT

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: Computer generated synthetic peptide

US-09-982-172-86

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 R 1

RESULT 17

US-09-982-172-93

; Sequence 93, Application US/09982172

; Patent No. US20020137119A1

; GENERAL INFORMATION:

; APPLICANT: Emil Israel Katz

; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES

; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING

; TITLE OF INVENTION: UTILIZING EACH

; FILE REFERENCE: 01/22283

; CURRENT APPLICATION NUMBER: US/09/982,172

; CURRENT FILING DATE: 2001-10-19

; NUMBER OF SEQ ID NOS: 253

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 93

; LENGTH: 1

; TYPE: PRT

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: Computer generated synthetic peptide

US-09-982-172-93

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 R 1

RESULT 18

US-09-982-172-93

; Sequence 93, Application US/09982172

; Patent No. US20020137119A1

; GENERAL INFORMATION:

; APPLICANT: Emil Israel Katz

; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES

; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING

; TITLE OF INVENTION: UTILIZING EACH

; FILE REFERENCE: 01/22283

; CURRENT APPLICATION NUMBER: US/09/982,172

; CURRENT FILING DATE: 2001-10-19

; NUMBER OF SEQ ID NOS: 253

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 93

; LENGTH: 1

; TYPE: PRT

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: Computer generated synthetic peptide

US-09-982-172-93

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 R 1


```
US-09-982-172-95
; Sequence 95, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 95
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-95
Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 X 1
Db 1 R 1
RESULT 19
US-09-982-172-106
; Sequence 106, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 106
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-106
Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 1; Gaps 0;
Qy 1 X 1
Db 1 K 1
RESULT 20
US-09-982-172-112
; Sequence 112, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 112
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-126
; Sequence 126, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 126
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-126
Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 1; Gaps 0;
Qy 1 X 1
Db 1 K 1
```

```
US-09-982-172-126
; Sequence 126, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 126
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-126
Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 1; Gaps 0;
Qy 1 X 1
Db 1 K 1
RESULT 21
US-09-982-172-120
; Sequence 120, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 120
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-120
Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 1; Gaps 0;
Qy 1 X 1
Db 1 K 1
RESULT 22
US-09-982-172-126
; Sequence 126, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 126
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-126
Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 1; Gaps 0;
Qy 1 X 1
Db 1 K 1
```


Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 1 R 1

RESULT 23
US-09-982-172-148
; Sequence 148, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 148
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-148

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 1 R 1

RESULT 24
US-09-982-172-149
; Sequence 149, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 149
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-149

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 1 Q 1

RESULT 25
US-09-982-172-155
; Sequence 155, Application US/09982172

; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 155
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-155

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 1 K 1

RESULT 26
US-09-982-172-160
; Sequence 160, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 160
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-160

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 1 K 1

RESULT 27
US-09-982-172-172
; Sequence 172, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 172
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-172

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 K 1

RESULT 28

US-09-982-172-173
; Sequence 173, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz

; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 173
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-173

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 R 1

RESULT 29

US-09-982-172-175
; Sequence 175, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz

; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 175
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-175

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 R 1

RESULT 30

US-09-982-172-189
; Sequence 189, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz

; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 189
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-189

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 K 1

RESULT 31

US-09-982-172-190
; Sequence 190, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz

; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 190
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-190

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 K 1

RESULT 32

US-09-982-172-191
; Sequence 191, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz

Query Match 0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;


```
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 191
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-191

Query Match          0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 K 1

RESULT 33
US-09-982-172-195
; Sequence 195, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 195
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-195

Query Match          0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 R 1

RESULT 34
US-09-982-172-200
; Sequence 200, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 200
; LENGTH: 1
```

```
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-200

Query Match          0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 K 1

RESULT 35
US-09-982-172-211
; Sequence 211, Application US/09982172
; Patent No. US20020137119A1
; GENERAL INFORMATION:
; APPLICANT: Emil Israel Katz
; TITLE OF INVENTION: PEPTIDES REPRESENTATIVE OF POLYPEPTIDES OF INTEREST AND ANTIBODIES
; TITLE OF INVENTION: DIRECTED THEREAGAINST, AND METHODS, SYSTEMS AND KITS FOR GENERATING
; TITLE OF INVENTION: UTILIZING EACH
; FILE REFERENCE: 01/22283
; CURRENT APPLICATION NUMBER: US/09/982,172
; CURRENT FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 211
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Computer generated synthetic peptide
US-09-982-172-211

Query Match          0.0%; Score 0; DB 9; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 K 1

RESULT 36
US-09-809-391-395
; Sequence 395, Application US/09809391
; Publication No. US20030049618A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P2
; CURRENT APPLICATION NUMBER: US/09/809,391
; CURRENT FILING DATE: 2001-03-16
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 761
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 395
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-809-391-395

Query Match          0.0%; Score 0; DB 10; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 M 1
```



```
RESULT 17
US-09-809-391-611
; Sequence 611, Application US/09809391
; Publication No. US20030049618A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P2
; CURRENT APPLICATION NUMBER: US/09/809,391
; CURRENT FILING DATE: 2001-03-16
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 761
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 611
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-809-391-611

Query Match      0.0%; Score 0; DB 10; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 X 1
Db      1 M 1

RESULT 38
US-09-882-171-395
; Sequence 395, Application US/09882171
; Publication No. US20030175858A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P2
; CURRENT APPLICATION NUMBER: US/09/882,171
; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 09/809,391
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 09/149,476
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: PCT/US98/04493
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,333
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/038,621
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,626
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,334
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,336
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,163
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/047,600
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,615
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,597
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,502
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,633
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,583
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,617
; PRIOR FILING DATE: 1997-05-23

; PRIOR APPLICATION NUMBER: 60/047,618
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,503
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,592
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,581
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,584
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,500
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,587
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,492
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,598
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,613
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,582
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,596
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,612
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,568
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,569
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,311
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,671
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,674
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,669
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,312
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,313
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,672
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,315
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/048,974
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/056,886
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,877
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,889
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,893
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,630
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,878
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,662
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,872
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,882
```


;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,637
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,903
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,888
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,879
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,880
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,894
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,911
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,636
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,874
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,910
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,864
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,631
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,845
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/047,595
;
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,599
;
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,588
;
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,585
;
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,586
;
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,590
;
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,594
;
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,589
;
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,593
;
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,614
;
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/043,578
;
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,576
;
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/047,501
;
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/043,670
;
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/056,632
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,664
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,876
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,881
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,909
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,875
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,862
;
; PRIOR FILING DATE: 1997-08-22

;
; PRIOR APPLICATION NUMBER: 60/056,887
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,908
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/048,964
;
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/057,650
;
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/056,884
;
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/057,669
;
; PRIOR FILING DATE: 1997-09-05

Query Match 0.0%; Score 0; DB 10; Length 1;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 M 1

RESULT 39
US-09-882-171-611
; Sequence 611, Application US/09882171
; Publication No. US20030175858A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P2
; CURRENT APPLICATION NUMBER: US/09/882,171
; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 09/809,391
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 09/149,476
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: PCT/US98/04493
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,333
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/038,621
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,626
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,334
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,336
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,163
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/047,600
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,615
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,597
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,502
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,633
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,583
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,617
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,618
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,503
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,592
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,581

[illegible]

↑
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/056,884
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/057,669
; PRIOR FILING DATE: 1997-09-05

Query Match 0.0%; Score 0; DB 10; Length 1;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1

Db 1 M 1

RESULT 40

US-09-833-245-184
; Sequence 184, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 184
; LENGTH: 1
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-245-184

Query Match 0.0%; Score 0; DB 11; Length 1;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1

Db 1 S 1

Search completed: April 13, 2005, 17:22:41
Job time : 8.32432 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2005, 17:02:44 ; Search time 2.34122 Seconds
(without alignments)
452.066 Million cell updates/sec

Title: US-09-998-861-14
Perfect score: 11
Sequence: 1 XXXXXXXXXX 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_79.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	0	0.0	3	3 RHST	thyroliberin - she
2	0	0.0	3	3 A92971	thyroliberin - eas
3	0	0.0	3	3 RHTDTC	thyroliberin - Bom
4	0	0.0	3	3 A22565	R-phycoerythrin al
5	0	0.0	3	3 PQ0010	angiotensin-conver
6	0	0.0	3	3 S13894	histidinol dehydro
7	0	0.0	3	3 A43391	TRH-like tripeptid
8	0	0.0	3	3 E37196	bradykinin-potenti
9	0	0.0	3	3 F37196	bradykinin-potenti
10	0	0.0	3	3 I50412	gene p20K protein
11	0	0.0	3	3 PT0636	T-cell receptor be
12	0	0.0	3	3 PT0578	T-cell receptor be
13	0	0.0	3	3 PT0571	T-cell receptor be
14	0	0.0	3	3 PT0622	T-cell receptor be
15	0	0.0	3	3 I78890	tyrosine protein k
16	0	0.0	3	3 S68328	blood cell protein
17	0	0.0	3	3 T13892	cytochrome-c oxida
18	0	0.0	3	3 GRHU	growth-modulating
19	0	0.0	3	3 RHPT	thyroliberin - pig
20	0	0.0	3	3 A60898	bursin - chicken
21	0	0.0	3	3 A23751	spinal cord peptid
22	0	0.0	3	3 B23751	spinal cord peptid
23	0	0.0	3	3 A33802	thyrotropin-releas
24	0	0.0	4	1 ECXAA	antho-RFamide neur
25	0	0.0	4	2 S18401	thyroglobulin - do
26	0	0.0	4	2 A02147	phagocytosis-stimu
27	0	0.0	4	2 A37832	phenol 2-monooxyge
28	0	0.0	4	2 A48360	gamma subunit of P
29	0	0.0	4	2 A61300	22K superhelical D

30	0	0.0	4	2 A41890	protein D - Becher
31	0	0.0	4	2 S43014	hypothetical prote
32	0	0.0	4	2 D41654	hypothetical prote
33	0	0.0	4	2 B43848	cell surface adhes
34	0	0.0	4	2 I40505	hypothetical prote
35	0	0.0	4	2 I40870	phospholipase C (E
36	0	0.0	4	2 T46627	hypothetical prote
37	0	0.0	4	2 S53508	starvation-induced
38	0	0.0	4	2 A27897	glucan 1,4-alpha-g
39	0	0.0	4	2 T30569	hypothetical prote
40	0	0.0	4	2 I38888	COI intron 16 prot
41	0	0.0	4	2 A25844	autho-RF amide neu
42	0	0.0	4	2 A34626	RPCH-related neuro
43	0	0.0	4	2 I51049	metallothionein-A
44	0	0.0	4	2 S39390	myosin-light-chain
45	0	0.0	4	2 I61883	protamine P1 - ora

ALIGNMENTS

RESULT 1

RHSHT

thyroliberin - sheep

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C>Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 09-Jul-2004

C:Accession: A93750; A01415

R:Desiderio Jr., D.M.; Burgus, R.; Dunn, T.F.; Vale, W.; Guillemin, R.; Ward, D.N.

Org. Mass Spectrom. 5, 221-228, 1971

A>Title: The elucidation of the primary structure of the hypothalamic thyrold stimulatir

A:Reference number: A93750

A:Accession: A93750

A:Molecule type: protein

A:Residues: 1-3 <DES>

A:Cross-references: UNIPROT:P01151

R:Burgus, R.; Dunn, T.F.; Desiderio, D.; Ward, D.N.; Vale, W.; Guillemin, R.

Nature 226, 321-325, 1970

A>Title: Characterization of ovine hypothalamic hypophysiotropic TSH-releasing factor.

A:Reference number: A93161; MUID:70163386; PMID:4985794

A:Contents: annotation

A>Note: physicochemical characteristics and biological activities of the natural and syn

C:Superfamily: thyroliberin

C:Keywords: amidated carboxyl end; hormone; hypothalamus; pyroglutamic acid

F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

F:3/Modified site: amidated carboxyl end (Pro) #status experimental

Query Match 0.0%; Score 0; DB 3; Length 3;

Best Local Similarity 0.0%; Pred.No. 0;

Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1

Db 1 Q 1

RESULT 2

A92971

thyroliberin - eastern newt (tentative sequence)

C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)

C>Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 09-Jul-2004

C:Accession: A92971; A01415

R:Grimm-Jorgensen, Y.; McKelvy, J.F.

J. Neurochem. 23, 471-478, 1974

A>Title: Biosynthesis of thyrotropin releasing factor by newt (Triturus viridescens) br

A:Reference number: A92971; MUID:75035605; PMID:4214528

A:Accession: A92971

A:Molecule type: protein

A:Residues: 1-3 <GRI>

A:Cross-references: UNIPROT:P01151

A>Note: a peptide with the chromatographic and electrophoretic characteristics of thyrol

istidine, or glutamic acid

C:Superfamily: thyroliberin

C:Keywords: amidated carboxyl end; cutaneous gland; hormone; hypothalamus; pyroglutamic

F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;3/Modified site: amidated carboxyl end (Pro) #status experimental

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 Q 1

RESULT 3

RHTDIO

C;Species: Bombina orientalis
C;Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 09-Jul-2004
C;Accession: A90919; A01415
R;Yasuhara, T.; Nakajima, T.
Chem. Pharm. Bull. 23, 3301-3303, 1975
A;Title: Occurrence of Pyl-His-Pro-NH-2 in the frog skin.
A;Reference number: A90919; MUID:76138399; PMID:815011
A;Accession: A90919

A;Molecule type: protein
A;Residues: 1-3 <YAS>
A;Cross-references: UNIPROT:P01151
C;Superfamily: thyroliberin
C;Keywords: amidated carboxyl end; cutaneous gland; hormone; pyroglutamic acid
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;3/Modified site: amidated carboxyl end (Pro) #status experimental

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 Q 1

RESULT 4

A22565

R-phycoerythrin alpha-1 chain - red alga (Gastrocloonium coulteri) (fragment)
C;Species: Gastrocloonium coulteri
C;Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Jun-2001
C;Accession: A22565
R;Klotz, A.V.; Glazer, A.N.
J. Biol. Chem. 260, 4856-4863, 1985
A;Title: Characterization of the bilin attachment sites in R-phycoerythrin.
A;Reference number: A22565; MUID:85182601; PMID:3886644
A;Accession: A22565

A;Molecule type: protein
A;Residues: 1-3 <KLO>

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 2 Y 2

RESULT 5

PQ0010

angiotensin-converting enzyme inhibitor (FLP-3) - common fig
N;Alternate names: ficus latex peptide 3
C;Species: Ficus carica (common fig)
C;Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Jun-2001
C;Accession: PQ0010
R;Maruyama, S.; Miyoshi, S.; Tanaka, H.
Agric. Biol. Chem. 53, 2763-2767, 1989
A;Title: Angiotensin I-converting enzyme inhibitors derived from Ficus carica.

A;Reference number: PQ0008
A;Accession: PQ0010
A;Molecule type: protein
A;Residues: 1-3 <MAR>
A;Experimental source: latex
C;Keywords: angiotensin-converting enzyme inhibitor

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 L 1

RESULT 6

S13894

histidinol dehydrogenase (EC 1.1.1.23) - wild cabbage (fragment)
C;Species: Brassica oleracea (wild cabbage)
C;Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Jun-2001
C;Accession: S13894
R;Nagai, A.; Scheidegger, A.
Arch. Biochem. Biophys. 284, 127-132, 1991
A;Title: Purification and characterization of histidinol dehydrogenase from cabbage.
A;Reference number: S13894; MUID:91112783; PMID:1989490
A;Accession: S13894
A;Molecule type: protein
A;Residues: 1-3 <NAG>
A;Experimental source: var. capitata
C;Keywords: dimer; NAD; oxidoreductase

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 X 1
DB 1 X 1

RESULT 7

A43391

TRH-like tripeptide - alfalfa
C;Species: Medicago sativa (alfalfa)
C;Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Jun-2001
C;Accession: A43391
R;Lackey, D.B.
J. Biol. Chem. 267, 17508-17511, 1992

A;Title: Isolation and structural determination of a novel TRH-like tripeptide, pyroglu-

A;Reference number: A43391; MUID:92388092; PMID:1517203
A;Accession: A43391

A;Molecule type: protein
A;Residues: 1-3 <IAC>
C;Keywords: amidated carboxyl end; pyroglutamic acid
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;3/Modified site: amidated carboxyl end (Pro) #status experimental

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 Q 1

RESULT 8

E37196

bradykinin-potentiating peptide 5 - island jararaca
C;Species: Bothrops insularis (island jararaca)
C;Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 30-Jun-2001
C;Accession: E37196

R;Cintra, A.C.O.; Vieira, C.A.; Giglio, J.R.

J. Protein Chem. 9, 221-227, 1990
A:Title: Primary structure and biological activity of bradykinin potentiating peptides
A:Reference number: A37196; MUID:90351557; PMID:2386615

C:Accession: E37196
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-3 <CIN>

A:Keywords: pyrroglutamic acid
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1

Db 1 Q 1

RESULT 9

F37196
bradykinin-potentiating peptide 6 - island jararaca
C:Species: Bothrops insularis (island jararaca)
C:Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Jun-2001

C:Accession: F37196
R;Cintra, A.C.O.; Vieira, C.A.; Giglio, J.R.

J. Protein Chem. 9, 221-227, 1990
A:Title: Primary structure and biological activity of bradykinin potentiating peptides
A:Reference number: A37196; MUID:90351557; PMID:2386615

C:Accession: F37196
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-3 <CIN>

A:Keywords: pyrroglutamic acid
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1

Db 1 Q 1

RESULT 10

I50412
gene p20K protein - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Jun-2001

C:Accession: I50412
R;Mao, P.L.; Beauchemin, M.; Bedard, P.A.

J. Biol. Chem. 268, 8131-8139, 1993
A:Title: Quiescence-dependent activation of the p20K promoter in growth-arrested chicken
A:Reference number: A46643; MUID:93216790; PMID:8463325

C:Accession: I50412
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-3 <MAO>

A:Cross-references: GB:L02537; NID:9212616; PID:9212617
C:Genetics:
A:Gene: p20K

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1

Db 2 S 2

RESULT 11

PT0636

T-cell receptor beta chain V-D-J region (100-2AT) - mouse (fragment)
C:Species: Mus musculus (house mouse)

C:Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Jun-2001
C:Accession: PT0636

R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991

A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A:Reference number: PT0509; MUID:91277601; PMID:1711558

C:Accession: PT0636
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-3 <PFE>

A:Experimental source: newborn thymus, strain BALB/c
C:Keywords: T-cell receptor

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1

Db 1 A 1

RESULT 12

PT0578

T-cell receptor beta chain V-D-J region (141-1BD) - mouse (fragment)
C:Species: Mus musculus (house mouse)

C:Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Jun-2001
C:Accession: PT0578

R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991

A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A:Reference number: PT0509; MUID:91277601; PMID:1711558

C:Accession: PT0578
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-3 <PFE>

A:Experimental source: day 19 fetal thymus, strain BALB/c
C:Keywords: T-cell receptor

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1

Db 1 A 1

RESULT 13

PT0571

T-cell receptor beta chain V-D-J region (141-1CM) - mouse (fragment)
C:Species: Mus musculus (house mouse)

C:Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Jun-2001
C:Accession: PT0571

R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991

A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A:Reference number: PT0509; MUID:91277601; PMID:1711558

C:Accession: PT0571
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-3 <PFE>

A:Experimental source: day 19 fetal thymus, strain BALB/c
C:Keywords: T-cell receptor

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 A 1

RESULT 14

PT0622
T-cell receptor beta chain V-D-J region - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Jun-2001
C:Accession: PT0622; PT0680; PT0582; PT0673
R:Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A:Reference number: PT0509; MUID:91277601; PMID:1711558
A:Accession: PT0622
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-3 <FEE>
A:Experimental source: newborn thymus, strain BALB/c, clone 111-1P
A:Accession: PT0680
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-3 <FEE1>
A:Experimental source: day 18 fetal thymus, strain BALB/c clones 154-1J and 154-1P
A:Accession: PT0582
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-3 <FEE2>
A:Experimental source: day 19 fetal thymus, strain BALB/c, clone 159-2A
C:Keywords: T-cell receptor

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 A 1

RESULT 15

I78890
tyrosine protein kinase - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Jun-2001
C:Accession: I78890
R:Chow, L.M.; Davidson, D.; Fournel, M.; Gosselin, P.; Lemieux, S.; Lyu, M.S.; Kozak, C.
Oncogene 9, 3437-3448, 1994
A:Title: Two distinct protein isoforms are encoded by ntk, a csk-related tyrosine protein kinase gene.
A:Reference number: I58407; MUID:95060800; PMID:7970703
A:Accession: I78890
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3 <RES>
A:Cross-references: GB:L33339; NID:G609536; PIDN:AAA64432.1; PID:G609538
C:Genetics:
A:Gene: p52ntk

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 3 T 3

RESULT 16

S68328
blood cell protein A - Molgula manhattensis (fragment)
C:Species: Molgula manhattensis
C:Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Jun-2001

C:Accession: S68328

R:Taylor, S.W.; Ross, M.M.; Waite, J.H.

Arch. Biochem. Biophys. 324, 228-240, 1995

A:Title: Novel 3,4-di- and 3,4,5-trihydroxyphenylalanine-containing polypeptides from the

A:Reference number: S68325; MUID:96132650; PMID:8554314

A:Accession: S68328

A:Molecule type: protein

A:Residues: 1-3 <TAY>

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 A 1

RESULT 17

T13892
cytochrome-c oxidase (EC 1.9.3.1) chain I [imported] - river lamprey mitochondrion (frag
C:Species: mitochondrion Lampetra fluviatilis (river lamprey)
C:Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Jun-2001
C:Accession: T13892
R:Delarbre, C.; Barriol, V.; Tillier, S.; Janvier, P.; Gachelin, G.
Mol. Biol. Evol. 14, 807-813, 1997
A:Title: The main features of the craniate mitochondrial DNA between the ND1 and the COI
A:Reference number: 217775; MUID:97398704; PMID:9254918
A:Accession: T13892
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3
A:Cross-references: EMBL:Y09528; NID:G2340016; PIDN:CAA70721.1; PID:G4379123
C:Genetics:
A:Genome: mitochondrion
A:Note: COI
C:Keywords: mitochondrion; oxidoreductase

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 2 T 2

RESULT 18

GKHU
growth-modulating peptide - human
C:Species: Homo sapiens (man)
C:Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Mar-2004
C:Accession: A01421
R:Schlesinger, D.H.; Pickart, L.; Thaler, M.M.
Experientia 33, 324-325, 1977
A:Title: Growth-modulating serum tripeptide is glycyl-histidyl-lysine.
A:Reference number: A01421; MUID:77162369; PMID:858356
A:Accession: A01421
A:Molecule type: protein
A:Residues: 1-3 <SCH>
A:Note: this serum tripeptide is found to stimulate growth of some cell types and to inhibit

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 G 1

RESULT 19

RHPGT

thyroliberin - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Mar-2004
C:Accession: A01415
R:Nair, R.M.G.; Barrett, J.F.; Bowers, C.Y.; Schally, A.V.
Biochemistry 9, 1103-1106, 1970
A:Title: Structure of porcine thyrotropin releasing hormone.
A:Reference number: A90560; MUID:70136150; PMID:4984938
A:Accession: A01415
A:Molecule type: protein
A:Residues: 1-3 <NAI>
R:Boiler, J.; Enmann, F.; Folkers, K.; Bowers, C.Y.; Schally, A.V.
Biochem. Biophys. Res. Commun. 37, 705-710, 1969
A:Title: The identity of chemical and hormonal properties of the thyrotropin releasing hormone.
A:Reference number: A90167; MUID:70039904; PMID:4982117
A:Contents: annotation
A>Note: biological activities and RF values (in 17 chromatographic systems) of the synthetic peptide: amidated carboxyl end; hormone; hypothalamus; pyroglutamic acid
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F:3/Modified site: amidated carboxyl end (Pro) #status experimental

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 Q 1

RESULT 20
A60898
bursin - chicken
C:Species: Gallus gallus (chicken)
C:Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Mar-2004
C:Accession: A60898
R:Audhya, T.; Kroon, D.; Heavner, G.; Viamontes, G.; Goldstein, G.
Science 231, 997-999, 1986
A:Title: Tripeptide structure of bursin, a selective B-cell-differentiating hormone of bursin.
A:Reference number: A60898; MUID:86122916; PMID:3484838
A:Accession: A60898
A:Molecule type: protein
A:Residues: 1-3 <AUD>
C:Keywords: amidated carboxyl end; hormone
F:3/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 K 1

RESULT 21
A23751
spinal cord peptide SCP-4 - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Mar-2004
C:Accession: A23751
R:Hsi, K.L.; Chen, R.L.; Chen, Z.G.; Zhang, H.L.; Lu, Y.A.; Guo, S.Y.; Wu, S.X.; Tsou, R.
Arch. Biochem. Biophys. 240, 178-183, 1985
A:Reference number: A23751; MUID:85250425; PMID:4015098
A:Accession: A23751
A>Status: Preliminary
A:Molecule type: protein
A:Residues: 1-3 <HSI>

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 K 1

RESULT 22
B23751
spinal cord peptide SCP-5 - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Mar-2004
C:Accession: B23751
R:Hsi, K.L.; Chen, R.L.; Chen, Z.G.; Zhang, H.L.; Lu, Y.A.; Guo, S.Y.; Wu, S.X.; Tsou, R.
Arch. Biochem. Biophys. 240, 178-183, 1985
A:Reference number: A23751; MUID:85250425; PMID:4015098
A:Accession: B23751
A>Status: Preliminary
A:Molecule type: protein
A:Residues: 1-3 <HSI>

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 M 1

RESULT 23
A33802
thyrotropin-releasing hormone-like peptide - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 15-Mar-2004
C:Accession: A33802
R:Cockle, S.M.; Aitken, A.; Beg, F.; Smyth, D.G.
J. Biol. Chem. 264, 7788-7791, 1989
A:Title: A novel peptide, pyroglutamylglutamylproline amide, in the rabbit prostate contains thyrotropin-releasing hormone-like peptide.
A:Reference number: A33802; MUID:89255196; PMID:2498305
A:Accession: A33802
A>Status: Preliminary
A:Molecule type: protein
A:Residues: 1-3 <COC>
C:Keywords: amidated carboxyl end; pyroglutamic acid
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F:3/Modified site: amidated carboxyl end (Pro) #status experimental

Query Match 0.0%; Score 0; DB 3; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 Q 1

RESULT 24
ECXAA
antho-RFamide neuropeptide - sea anemone (Anthopleura elegantissima)
C:Species: Anthopleura elegantissima
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jul-2004
C:Accession: A26666
R:Grimmelikhuizen, C.J.P.; Graff, D.
Proc. Natl. Acad. Sci. U.S.A. 83, 9817-9821, 1986
A:Title: Isolation of <Glu-Gly-Arg-Phe-NH2 (Antho-RFamide), a neuropeptide from sea anemone.
A:Reference number: A26666; MUID:87092339; PMID:2879288
A:Accession: A26666
A:Molecule type: protein
A:Residues: 1-4 <GRI>
A:Cross-references: UNIPROT:P10419
C:Comment: The function of this peptide is not known but it could act as a transmitter.
C:Comment: Synthetic and natural peptides had identical properties.
C:Superfamily: RFamide neuropeptide
C:Keywords: amidated carboxyl end; neuropeptide; pyroglutamic acid

F,1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F,4/Modified site: amidated carboxyl end (Phe) #status experimental

Query Match 0.0%; Score 0; DB 1; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 Q 1

RESULT 25

SI8401
thyroglobulin - dog (fragment)
C,Species: Canis lupus familiaris (dog)
C,Date: 19-Mar-1997 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C,Accession: SI8401
R,Donada, A.; Vassart, G.; Christophe, D.
Biochim. Biophys. Acta 1090, 235-237, 1991
A,Title: Isolation and characterization of the canine thyroglobulin gene promoter region
A,Reference number: SI8401; MUID:92031697; PMID:1932116
A,Accession: SI8401
A,Status: preliminary
A, Molecule type: DNA
A,Residues: 1-4 <DON>
A,Cross-references: GB:S61184; NID:9237714; PIDN:AAB20127.1; PID:9237715
C,Superfamily: thyroglobulin; cholinesterase homology; thyroglobulin type I repeat homol
C,Keywords: duplication; iodine; thyroid gland; thyroid hormone biosynthesis

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 2 A 2

RESULT 26

A02147
phagocytosis-stimulating peptide (tuftsin) - human
C,Species: Homo sapiens (man)
C,Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 09-Jul-2004
C,Accession: A02147
R,Nishioaka, K.; Constantopoulos, A.; Satoh, P.S.; Najjar, V.A.
Biochem. Biophys. Res Commun. 47, 172-179, 1972
A,Title: The characteristics, isolation and synthesis of the phagocytosis stimulating pe
A,Reference number: A02147; MUID:72187087; PMID:4112769
A,Accession: A02147
A, Molecule type: protein
A,Residues: 1-4 <NIS>
A,Cross-references: UNIPROT:P01858
A,Note: a peptide having the same structure, physical properties, and biological activit
R,Fidalgo, B.V.; Najjar, V.A.
Biochemistry 6, 3386-3392, 1967
A,Reference number: A37502; MUID:68091045; PMID:4169272
A,Contents: annotation; immunoglobulin class
C,Comment: An IgG (called leucokinin) binds reversibly to the cell membrane of neutrophil
N is essential for maximum stimulation of the phagocytic activity of neutrophils.
C,Superfamily: immunoglobulin C region; immunoglobulin homology

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 T 1

RESULT 27

A37832

phenol 2-monooxygenase (BC 1.14.13.7) chain P5 - Pseudomonas sp. (strain CF600) (fragment)
C,Species: Pseudomonas sp.
C,Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 23-Jun-1993
C,Accession: A37832
R,Powlowski, J.; Shingler, V.
J. Bacteriol. 172, 6834-6840, 1990
A,Title: In vitro analysis of polypeptide requirements of multicomponent phenol hydroxyl
A,Reference number: A37832; MUID:91072231; PMID:2254259
A,Accession: A37832
A,Status: preliminary
A, Molecule type: protein
A,Residues: 1-4 <POW>
C,Keywords: oxidoreductase

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 S 1

RESULT 28

A48360
gamma subunit of Protein A - Methylosinus trichosporium (fragment)
C,Species: Methylosinus trichosporium
C,Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C,Accession: A48360
R,Cardy, D.B.; Laidler, V.; Salmon, G.P.; Murrell, J.C.
Arch. Microbiol. 156, 477-483, 1991
A,Title: The methane monooxygenase gene cluster of Methylosinus trichosporium: cloning a
A,Reference number: A48360; MUID:92153031; PMID:1785954
A,Contents: OB3b
A,Accession: A48360
A,Status: preliminary
A, Molecule type: DNA
A,Residues: 1-4 <CAR>
A,Cross-references: GB:S81887; NID:9245213; PIDN:AAB21391.1; PID:9245214
A,Note: sequence extracted from NCBI backbone (NCBI:81887, NCBIP:81912)

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 4 A 4

RESULT 29

A61300
22K superhelical DNA-binding protein - Escherichia coli (fragment)
C,Species: Escherichia coli
C,Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 07-May-1999
C,Accession: A61300
R,Kishi, F.; Ebina, Y.; Miki, T.; Nakazawa, T.; Nakazawa, A.
J. Biochem. 92, 1059-1068, 1982
A,Title: Purification and characterization of a protein from Escherichia coli which form
A,Reference number: A61300; MUID:83082696; PMID:6294066
A,Accession: A61300
A, Molecule type: protein
A,Residues: 1-4 <KIS>
C,Comment: This protein resembles some of the histone-like protein of bacteria in amino
C,Keywords: DNA binding; monomer

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 M 1

RESULT 30
A41890
protein D - Escherichia coli (fragment)
C:Species: Escherichia coli
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993
C:Accession: A41890
R:Slottan, A.; Gebhardt, K.; Kristiansen, E.; Birkeland, N.K.; Lindqvist, B.H.
J. Bacteriol. 174, 4094-4100, 1992
A:Title: Escherichia coli K-12 and B contain functional bacteriophage P2 ogr genes.
A:Reference number: A41890; MUID:92283767; PMID:1597424
A:Accession: A41890
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4 <SLE>
A:Cross-references: GB:M81463

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 E 1

RESULT 31
S43014
hypothetical protein URF-2Y - Yersinia enterocolitica transposon TN3926
C:Species: Yersinia enterocolitica
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
C:Accession: S43014
R:Osbourne, S.E.V.; Turner, A.K.; Grinstead, J.
submitted to the EMBL Data Library, March 1994
A:Description: The structure of the bacterial transposable element, Tn3926.
A:Reference number: S43011
A:Accession: S43014
A:Molecule type: DNA
A:Residues: 1-4 <OSB>
A:Cross-references: EMBL:X78059; NID:9460067; PIDN:CAA54980.1; PID:9581836
C:Genetics:
A:Mobile element: transposon TN3926

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 3 A 3

RESULT 32
D41654
hypothetical protein (sodC 5' region) - Haemophilus parainfluenzae (fragment)
C:Species: Haemophilus parainfluenzae
C:Date: 12-Jun-1992 #sequence_revision 12-Jun-1992 #text_change 24-Feb-1995
C:Accession: D41654
R:Kroll, J.S.; Langford, P.R.; Loynds, B.M.
J. Bacteriol. 173, 7449-7457, 1991
A:Title: Copper-zinc superoxide dismutase of Haemophilus influenzae and Haemophilus para
A:Reference number: D41654; MUID:92041655; PMID:1938942
A:Accession: D41654
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4 <KRO>

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 E 1

Db 1 D 1

RESULT 33
B43848
cell surface adhesin for heparan sulfate, 60K - Staphylococcus aureus (fragment)
C:Species: Staphylococcus aureus
C:Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 24-Feb-1995
C:Accession: B43848
R:Liang, O.D.; Ascencio, F.; Fransson, L.A.; Wadstrom, T.
Infect. Immun. 60, 899-906, 1992
A:Title: Binding of heparan sulfate to Staphylococcus aureus.
A:Reference number: A43848; MUID:92176005; PMID:1541563
A:Accession: B43848
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-4 <LIA>
A:Note: sequence extracted from NCBI backbone (NCBI:85444)

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 4 T 4

RESULT 34
I40505
hypothetical protein 3 (4 aa) - Bacillus stearothermophilus
C:Species: Bacillus stearothermophilus
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Oct-1999
C:Accession: I40505
R:Waye, M.M.; Winter, G.
Eur. J. Biochem. 158, 505-510, 1986
A:Title: A transcription terminator in the 5' non-coding region of the tyrosyl tRNA synt
A:Reference number: I40503; MUID:86274732; PMID:3525162
A:Accession: I40505
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-4 <RES>
A:Cross-references: EMBL:X04193; NID:940233; PIDN:CAA27783.1; PID:9580944

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 3 S 3

RESULT 35
I40870
phospholipase C (EC 3.1.4.3) - Clostridium perfringens (fragment)
C:Species: Clostridium perfringens
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 21-Jul-2000
C:Accession: I40870
R:Toyonaga, T.; Matsushita, O.; Katayama, S.; Minami, J.; Okabe, A.
Microbiol. Immunol. 36, 603-613, 1992
A:Title: Role of the upstream region containing an intrinsic DNA curvature in the negati
A:Reference number: I40870; MUID:92396045; PMID:1522810
A:Accession: I40870
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-4 <RES>
A:Cross-references: EMBL:X62825; NID:940622; PIDN:CAA44636.1; PID:94377417
C:Genetics:
A:Gene: plc
C:Keywords: phosphoric diester hydrolase

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 M 1

RESULT 36
T46627
hypothetical protein c4 - loblolly pine
C:Species: Pinus taeda (loblolly pine)
C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 18-Feb-2000
C:Accession: T46627
R:Chang, S.; Puryea, J.; Funkhouser, E.A.; Newton, R.J.; Cairney, J.
submitted to the EMBL Data Library, July 1995
A:Description: Cloning of a chitinase homolog which lacks chitin binding sites and is do
A:Reference number: Z23105
A:Accession: T46627
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4 <CHA>
A:Cross-references: EMBL:U31309; NID:G974285; PID:G974292
A:Experimental source: strain s6PT2xs6PT3; 8 month seedlings

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 M 1

RESULT 37
S53508
starvation-induced ribonuclease - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 07-May-1999
C:Accession: S53508
R:Koeck, M.; Loeffler, A.; Abel, S.; Glund, K.
Plant Mol. Biol. 27, 477-485, 1995
A:Title: cDNA structure and regulatory properties of a family of starvation-induced ribo
A:Reference number: S53506; MUID:95201242; PMID:7894013
A:Accession: S53508
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-4 <KOE>

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 3 T 3

RESULT 38
A27897
Glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - Aspergillus phoenicis (fragment)
N:Alternate names: glucoamylase
C:Species: Aspergillus phoenicis
C:Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 06-Dec-1996
C:Accession: A27897
R:Inokuchi, N.; Takahashi, T.; Irie, M.
J. Biochem. 90, 1055-1067, 1981
A:Title: Purification and characterization of a minor glucoamylase from Aspergillus said
A:Reference number: A27897; MUID:82075730; PMID:6796572
A>Note: Aspergillus saitoi
A:Accession: A27897
A:Molecule type: protein

A:Residues: 1-4 <INO>
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 A 1

RESULT 39
T30569
hypothetical protein - Emericella nidulans
C:Species: Emericella nidulans, Aspergillus nidulans
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C:Accession: T30569
R:Morrice, J.; MacKenzie, D.A.; Parr, A.J.; Archer, D.B.
Curr. Genet. 34, 379-385, 1998
A:Title: Isolation and characterisation of the acetyl-CoA carboxylase gene from Aspergil
A:Reference number: Z20869; MUID:99087906; PMID:9871120
A:Accession: T30569
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4 <MOR>
A:Cross-references: EMBL:Y15996; NID:e1285512; PID:e1218041; PIDN:CAA75927.1

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 3 S 3

RESULT 40
I38888
COI intron 16 protein - Podospora anserina mitochondrion
C:Species: mitochondrion Podospora anserina
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 07-Dec-1999
C:Accession: I38888
R:Cummings, D.J.; Michel, F.; McNally, K.L.
Curr. Genet. 16, 381-406, 1989
A:Title: DNA sequence analysis of the 24.5 kilobase pair cytochrome oxidase subunit I m
A:Reference number: A48327; MUID:90124722; PMID:2558809
A:Accession: I38888
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4 <CUM>
A:Cross-references: GB:X55026; GB:M30937; GB:M61734
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC3
C:Keywords: mitochondrion

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 X 1
Db 1 Q 1

Search completed: April 13, 2005, 17:17:29
Job time : 3.34122 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2005, 17:01:49 ; Search time 10.5912 Seconds
(without alignments)
531.844 Million cell updates/sec

Title: US-09-998-861-14
Perfect score: 11
Sequence: 1 XXXXXXXXXXXX 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	0	0.0	1 GWA SEPOF	P83570 sepia offic
2	0	0.0	3 1 GRWM HUMAN	P01157 homo sapien
3	0	0.0	3 1 LUXE VIBFI	P24272 vibrio fisci
4	0	0.0	3 1 THYL BOMOR	P62970 bombina ori
5	0	0.0	3 1 THYL NOTVI	P62971 notophthalm
6	0	0.0	3 1 THYL PIG	P62968 sus scrofa
7	0	0.0	3 1 THYL SHEEP	P62969 ovis aries
8	0	0.0	4 1 ACHI ACHFU	P35904 achatina fu
9	0	0.0	4 1 DCML PSECH	P19916 pseudomonas
10	0	0.0	4 1 DCMS PSECH	P19918 pseudomonas
11	0	0.0	4 1 EOSI HUMAN	P02731 homo sapien
12	0	0.0	4 1 FAR3 HIRME	P42562 hirudo medi
13	0	0.0	4 1 FAR4 HIRME	P42563 hirudo medi
14	0	0.0	4 1 FFKA ANTEL	P58705 anthopleura
15	0	0.0	4 1 FLRF HIRME	P42561 hirudo medi
16	0	0.0	4 1 FLRN ANTEL	P58707 anthopleura
17	0	0.0	4 1 FMRF MACNI	P01162 macrocallis
18	0	0.0	4 1 FYRI ANTEL	P58706 anthopleura
19	0	0.0	4 1 ILME SEPOF	P83568 sepia offic
20	0	0.0	4 1 OCP1 OCTMI	P58648 octopus min
21	0	0.0	4 1 OCP3 OCTMI	P58649 octopus min
22	0	0.0	4 1 TUFT HUMAN	P01858 homo sapien
23	0	0.0	4 1 YLM1 YEAST	P36515 saccharomyc
24	0	0.0	4 2 Q16047	Q16047 homo sapien
25	0	0.0	4 2 Q98A70	Q98A70 homo sapien
26	0	0.0	4 2 Q08433	Q08433 rattus sp.
27	0	0.0	5 1 AL14 CARMA	P81817 carcinus ma
28	0	0.0	5 1 AP21 BISFO	P84182 eisenia foe
29	0	0.0	5 1 BIOA CITFR	P13071 citrobacter
30	0	0.0	5 1 BIOB CITFR	P12997 citrobacter
31	0	0.0	5 1 BP77 BOTIN	P30425 botriopsis in

32	0	0.0	5 1 EI03 LITRU	P82099 litoria rub
33	0	0.0	5 1 EI04 LITRU	P82100 litoria rub
34	0	0.0	5 1 FARP ARTTR	P41853 artiopesthi
35	0	0.0	5 1 FARP CHICK	P83308 gallus gall
36	0	0.0	5 1 MPA4 JUNVI	P81826 juniperus v
37	0	0.0	5 1 PAP2 PARMA	P81864 pardachirus
38	0	0.0	5 1 PRCT CARMA	P67857 carcinus ma
39	0	0.0	5 1 PRCT LIMPO	P67858 limulus pol
40	0	0.0	5 1 PRCT PERAM	P67859 periplaneta
41	0	0.0	5 1 PSK DAUCA	P58261 daucus caro
42	0	0.0	5 1 RE11 LITRU	P82070 litoria rub
43	0	0.0	5 1 RE21 LITRU	P82071 litoria rub
44	0	0.0	5 1 RE31 LITRU	P82072 litoria rub
45	0	0.0	5 1 RE32 LITRU	P82073 litoria rub

ALIGNMENTS

RESULT 1
GWA SEPOF STANDARD; PRT; 2 AA.
AC P83570;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Neuropeptide GWA.
OS Sepia officinalis (Common cuttlefish).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
OC Decapodiformes; Sepioidea; Sepiidae; Sepia.
OX NCBI_TaxID=6610;
RN [1]
RP SEQUENCE, FUNCTION, MASS SPECTROMETRY, AND AMIDATION.
RC TISSUE=Optic lobe;
RX MEDLINE=99100358; PubMed=9437704; DOI=10.1016/S0196-9781(97)00241-6;
RA Henry J., Favrel P., Boucaud-Camou E.;
RT "Isolation and identification of a novel Ala-Pro-Gly-Trp-amide-related
RT peptide inhibiting the motility of the mature oviduct in the
RT cuttlefish, Sepia officinalis.";
RL Peptides 18:1469-1474(1997).
CC -!- FUNCTION: Regulatory neuropeptide with myotropic activity
CC targeting the distal oviduct. Inhibits the motility of the oviduct
CC by decreasing tonus, frequency and amplitude of contractions.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MASS SPECTROMETRY: MW=259.9; METHOD=WALDI; RANGE=1-2; NOTE=Ref.1.
KW Amidation; Direct protein sequencing; Neuropeptide.
FT MOD RES 2 Tryptophan amide.
SQ SEQUENCE 2 AA; 261 MW; 7378100000000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 2;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 X 1

Db 1 G 1

RESULT 2

GRWM HUMAN STANDARD; PRT; 3 AA.
ID GRWM HUMAN
AC P01157;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Growth-modulating peptide.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=77162369; PubMed=858356;


```

RA Schlesinger D.H., Pickart L., Thaler M.M.;
RT "Growth-modulating serum tripeptide is Glycyl-histidyl-lysine.";
RL Experimentia 33:324-325(1977).
CC -|- MISCELLANEOUS: This serum tripeptide has been found to stimulate
CC growth of some cell types and to inhibit other types in vitro.
DR GO; GO:0001556; P:regulation of cell growth; NAS.
KW Direct protein sequencing.
SQ SEQUENCE 3 AA; 340 MW; 6331E8100000000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 3;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 G 1

RESULT 3
LUXE_VIBFI STANDARD; PRT; 3 AA.
AC P24572;
DT 01-WAR-1992 (Rel. 21, Created)
DT 01-WAR-1992 (Rel. 21, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Long-chain-fatty-acid--luciferin-component ligase (EC 6.2.1.19) (Acyl-
DE protein synthetase) (Fragment).
GN Name=luxE;
OS Vibrio fischeri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91072226; PubMed=2254256;
RA Swartzman E., Kapoor S., Graham A.F., Meighen E.A.;
RT "A new Vibrio fischeri lux gene precedes a bidirectional termination
RT site for the lux operon.";
RL J. Bacteriol. 172:6797-6802(1990).
CC -|- FUNCTION: Acyl-protein synthetase activates tetradecanoic acid. It
CC is a component of the fatty acid reductase complex responsible for
CC converting tetradecanoic acid to the aldehyde which serves as
CC substrate in the luciferase-catalyzed reaction.
CC -|- CATALYTIC ACTIVITY: ATP + an acid + protein = AMP + diphosphate +
CC an acyl-protein thioester.
CC -|- PATHWAY: Bioluminescent fatty acid reduction system; second step.
CC -|- SIMILARITY: Belongs to the luxE family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M62812; -; NOT ANNOTATED_CDS.
KW Ligase; Luminescence.
FT NON_TER 1
SQ SEQUENCE 3 AA; 374 MW; 6AA3303000000000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 3;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 I 1

RESULT 4
THYL_BOMOR STANDARD; PRT; 3 AA.

```

```

AC P62970; P01151;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Thyroliberin (Thyrotropin-releasing hormone) (TRH) (Thyrotropin
DE releasing factor) (TSH-releasing factor) (Protirelin).
OS Bombina orientalis (Oriental fire-bellied toad).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Archeobatrachia; Bombinatoridae; Bombina.
OX NCBI_TaxID=8346;
RN [1]
RP SEQUENCE.
RC TISSUE=Skin;
RX MEDLINE=76138399; PubMed=815011;
RA Yasuhara T., Nakajima T.;
RT "Occurrence of Pyr-His-Pro-NH2 in the frog skin.";
RL Chem. Pharm. Bull. 23:3301-3303(1975).
CC -|- FUNCTION: TRH functions as a regulator of the biosynthesis of TSH
CC in the anterior pituitary gland and as a neurotransmitter/
CC neuromodulator in the central and peripheral nervous systems (By
CC similarity).
CC -|- SUBCELLULAR LOCATION: Secreted.
DR PIR; A90919; RHDT0.
KW Amidation; Direct protein sequencing; Pyrrolidone carboxylic acid.
FT MOD_RES 1 1 Pyrrolidone carboxylic acid.
FT MOD_RES 3 3 Proline amide.
SQ SEQUENCE 3 AA; 380 MW; 7761F6B00000000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 3;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
DB 1 Q 1

RESULT 5
THYL_NOTVI STANDARD; PRT; 3 AA.
AC P62971; P01151;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Thyroliberin (Thyrotropin-releasing hormone) (TRH) (Thyrotropin
DE releasing factor) (TSH-releasing factor) (Protirelin).
OS Notoptthalmus viridescens (Eastern newt) (Triturus viridescens);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae;
OX Notoptthalmus.
OX NCBI_TaxID=8316;
RN [1]
RP SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=75035605; PubMed=4214528;
RA Grimm-Joergensen Y., McKelvy J.F.;
RT "Biosynthesis of thyrotropin releasing factor by newt (Triturus
RT viridescens) brain in vitro. Isolation and characterization of
RT thyrotropin releasing factor.";
RL J. Neurochem. 23:471-478(1974).
CC -|- FUNCTION: TRH functions as a regulator of the biosynthesis of TSH
CC in the anterior pituitary gland and as a neurotransmitter/
CC neuromodulator in the central and peripheral nervous systems.
CC -|- SUBCELLULAR LOCATION: Secreted.
DR PIR; A92971; A92971.
KW Amidation; Direct protein sequencing; Pyrrolidone carboxylic acid.
FT MOD_RES 1 1 Pyrrolidone carboxylic acid.
FT MOD_RES 3 3 Proline amide.
SQ SEQUENCE 3 AA; 380 MW; 7761F6B00000000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 3;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```


QY 1 X 1
Db 1 Q 1

RESULT 6

THYL_PIG STANDARD; PRT; 3 AA.
AC P62968; P01151;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Thyroliberin (Thyrotropin releasing hormone) (TRH) (Thyrotropin releasing factor) (TSH-releasing factor) (Protirelin).
GN Name=TRH;
OS Sus scrofa. (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RN [2]
RP SEQUENCE.
RC TISSUE=Hypothalamus;
RX MEDLINE=70136150; PubMed=4984938;
RA Nair R.M.G., Barrett J.F., Bowers C.Y., Schally A.V.;
RT "Structure of porcine thyrotropin releasing hormone.";
RL Biochemistry 9:1103-1106 (1970).
RN [2]
RP SYNTHESIS.
RX MEDLINE=70039904; PubMed=4982117;
RA Boler J., Enzmann F., Folkers K., Bowers C.Y., Schally A.V.;
RT "The identity of chemical and hormonal properties of the thyrotropin releasing hormone and pyroglutamyl-histidyl-proline amide.";
RL Biochem. Biophys. Res. Commun. 37:705-710 (1969).
CC -!- FUNCTION: TRH functions as a regulator of the biosynthesis of TSH in the anterior pituitary gland and as a neurotransmitter/neuromodulator in the central and peripheral nervous systems.
CC -!- SUBCELLULAR LOCATION: Secreted.
KW Amidation; Direct protein sequencing; Pyrrolidone carboxylic acid.
FT MOD_RES 1 1 Pyrrolidone carboxylic acid.
FT MOD_RES 3 3 Proline amide.
SQ SEQUENCE 3 AA; 380 MW; 7761F6B000000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 1 Q 1

RESULT 7

THYL_SHEEP STANDARD; PRT; 3 AA.
AC P62969; P01151;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Thyroliberin (Thyrotropin releasing hormone) (TRH) (Thyrotropin releasing factor) (TSH-releasing factor) (Protirelin).
GN Name=TRH;
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; OC Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RN [2]
RP SEQUENCE.
RC TISSUE=Hypothalamus;
RX Desiderio D.M. Jr., Burgess R., Dunn T.F., Vale W., Guillemin R.;
RA Ward D.N.;
RT "The elucidation of the primary structure of the hypothalamic thyroid

RT stimulating hormone releasing factor of ovine origin by means of mass spectrometry.";
RL Org. Mass Spectrom. 5:221-228 (1971).
RN [2]
RP SYNTHESIS.
RX MEDLINE=70163386; PubMed=4985794;
RA Burgess R., Dunn T.F., Desiderio D.M., Ward D.N., Vale W.,
RA Guillemin R.;
RT "Characterization of ovine hypothalamic hypophysiotropic TSH-releasing factor.";
RL Nature 226:321-325 (1970).
CC -!- FUNCTION: TRH functions as a regulator of the biosynthesis of TSH in the anterior pituitary gland and as a neurotransmitter/neuromodulator in the central and peripheral nervous systems.
CC -!- SUBCELLULAR LOCATION: Secreted.
KW PIR; A93750; RHSH.
DR Amidation; Direct protein sequencing; Pyrrolidone carboxylic acid.
FT MOD_RES 1 1 Pyrrolidone carboxylic acid.
FT MOD_RES 3 3 Proline amide.
SQ SEQUENCE 3 AA; 380 MW; 7761F6B000000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 3;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 1 Q 1

RESULT 8

ACHIL_ACHFU STANDARD; PRT; 4 AA.
AC P35904;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Achatin-I.
OS Achatina fulica (Giant African snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora; OC Sigmurethra; Achatinoidea; Achatinidae; Achatina.
OX NCBI_TaxID=6530;
RN [1]
RP SEQUENCE, CHARACTERIZATION, AND SYNTHESIS.
RC STRAIN=Ferussac; TISSUE=Ganglion;
RX MEDLINE=89273551; PubMed=2597281;
RA Kamatani Y., Minakata H., Kenny P.T.M., Iwashita T., Watanabe K.,
RA Funase K., Sun X.P., Yongsiri A., Kim K.H., Novales-Li P.,
RA Novales E.T., Kanapi C.G., Takeuchi H., Nomoto K.;
RT "Achatin-I, an endogenous neuroexcitatory tetrapeptide from Achatina fulica Ferussac containing a D-amino acid residue.";
RL Biochem. Biophys. Res. Commun. 160:1015-1020 (1989).
RN [2]
RP CHARACTERIZATION.
RC STRAIN=Ferussac; TISSUE=Heart atrium;
RX MEDLINE=91264856; PubMed=1675568;
RA Fujimoto K., Kubota I., Yasuda-Kamatani Y., Minakata H., Nomoto K.,
RA Yoshida M., Harada A., Muneoka Y., Kobayashi M.;
RT "Purification of achatin-I from the atria of the African giant snail, Achatina fulica, and its possible function.";
RL Biochem. Biophys. Res. Commun. 177:847-853 (1991).
RN [3]
RP CRYSTALLIZATION.
RX MEDLINE=93014529; PubMed=1399265;
RA Ishida T., In Y., Doi M., Inoue M., Yasuda-Kamatani Y., Minakata H.,
RA Iwashita T., Nomoto K.;
RT "Crystal structure and molecular conformation of achatin-I (H-Gly-D-Phe-Ala-Asp-OH), an endogenous neuropeptide containing a D-amino acid residue.";
RL Int. J. Pept. Protein Res. 39:258-264 (1992).
CC -!- FUNCTION: Neuroexcitatory peptide; increases the impulse frequency and produces a spike broadening of the identified heart excitatory neuron (PON); also enhances the amplitude and frequency of the

CC heart beat. Has also an effect on several other muscies.
 DR PIR; A32480; A32480.
 KW D-amino acid; Direct protein sequencing; Hormone.
 FT MOD_RES 2 2
 SQ SEQUENCE 4 AA; 408 MW; 6AADD9C810000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 4;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1

Db 3 A 3

RESULT 9

DCML_PSECH STANDARD; PRT; 4 AA.
 AC P19316;

DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Carbon monoxide dehydrogenase large chain (EC 1.2.99.2) (CO
 dehydrogenase subunit L) (CO-DH L) (Fragment).
 GN Name=cutL;

OS Pseudomonas carboxydohydrogena.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae.
 OX NCBI_TaxID=290;

RN [1]

RP SEQUENCE.

RX MEDLINE=90055678; PubMed=2818128;

RA Kraut M., Hugendieck I., Herwig S., Meyer O.;

RT "Homology and distribution of CO dehydrogenase structural genes in

carboxydohydrogenic bacteria.";

RL Arch. Microbiol. 152:335-341 (1989).

CC -!- FUNCTION: Catalyzes the oxidation of carbon monoxide to carbon
 dioxide.

CC -!- CATALYTIC ACTIVITY: CO + H(2)O + A = CO(2) + AH(2).

CC -!- COFACTOR: Binds 1 copper(I) ion, 1 molybdenum(VI) ion and 1

molybdopterin cytosine dinucleotide (MCD) per subunit.

CC -!- SUBUNIT: Heterotrimer consisting of a large, a medium and a small
 subunit.

DR PIR; P19140; P19140.

KW Direct protein sequencing; Molybdenum; Oxidoreductase.

FT NON_TER 4 4

SQ SEQUENCE 4 AA; 441 MW; 7761E876F0000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 4;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1

Db 1 M 1

RESULT 10

DCMS_PSECH

ID DCMS_PSECH STANDARD; PRT; 4 AA.
 AC P19318;

DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Carbon monoxide dehydrogenase small chain (EC 1.2.99.2) (CO
 dehydrogenase subunit S) (CO-DH S) (Fragment).
 GN Name=cutS;

OS Pseudomonas carboxydohydrogena.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Bradyrhizobiaceae.

OX NCBI_TaxID=290;

RN [1]

RP SEQUENCE.

RX MEDLINE=90055678; PubMed=2818128;
 RA Kraut M., Hugendieck I., Herwig S., Meyer O.;

RT "Homology and distribution of CO dehydrogenase structural genes in
 carboxydohydrogenic bacteria.";

RL Arch. Microbiol. 152:335-341 (1989).

CC -!- FUNCTION: Catalyzes the oxidation of carbon monoxide to carbon
 dioxide.

CC -!- CATALYTIC ACTIVITY: CO + H(2)O + A = CO(2) + AH(2).

CC -!- COFACTOR: Binds 2 2Fe-2S clusters (By similarity).

CC -!- SUBUNIT: Heterotrimer consisting of a large, a medium and a small
 subunit.

DR PIR; P19146; P19146.

KW 2Fe-2S; Direct protein sequencing; Iron; Iron-sulfur; Metal-binding;
 Oxidoreductase.

FT NON_TER 4 4

SQ SEQUENCE 4 AA; 420 MW; 6DD33DD6F0000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 4;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1

Db 2 A 2

RESULT 11

EOSI_HUMAN

ID EOSI_HUMAN STANDARD; PRT; 4 AA.
 AC P02731;

DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Eosinophilotoxic peptides.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE.

RX MEDLINE=76078412; PubMed=1060093;

RA Goetzl E.J., Austen K.F.;

RT "Purification and synthesis of eosinophilotoxic tetrapeptides of
 human lung tissue: identification as eosinophil chemotactic factor of
 anaphylaxis.";

RL Proc. Natl. Acad. Sci. U.S.A. 72:4123-4127 (1975).

CC -!- MISCELLANEOUS: These peptides are released from mast cells in lung
 (and other tissues) during hypersensitivity reactions
 (anaphylaxis). Their activities, preferentially affecting
 eosinophils, include chemotaxis, chemotactic deactivation, release
 of enzymes, and stimulation of the hexose monophosphate shunt.

CC GO:0006935; P:chemotaxis; IDA.

DR GO:0006935; P:chemotaxis; IDA.

DR GO:0006935; P:immune response; IDA.

KW Direct protein sequencing.

FT VARIANT 1 1

FT V -> A (in other peptide).
 /FTID=VAR_005201.

SQ SEQUENCE 4 AA; 390 MW; 6B05B862A0000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 4;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1

Db 3 S 3

RESULT 12

FAR3_HIRME

ID FAR3_HIRME STANDARD; PRT; 4 AA.
 AC P42562;

DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE FMRamide-like neuropeptide YLRF-amide.
 OS Hirudo medicinalis (Medicinal leech).
 OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinidae; Hirudo.
 OX NCBI_TaxID=6421;
 RN [1]
 RP SEQUENCE.
 RA MEDLINE=92195954; PubMed=1686933; DOI=10.1016/0196-9781(91)90035-N;
 RX Evans B.D., Pohl J., Kartsonis M.A., Calabrese R.L.;
 RT "Identification of Rfamide neuropeptides in the medicinal leech.";
 RL Peptides 12:897-908(1991).
 CC -|- SUBCELLULAR LOCATION: Secreted.
 CC -|- SIMILARITY: Belongs to the FARP (FMRamide related peptide) family.
 KW Amidation; Direct protein sequencing; Neuropeptide.
 FT MOD RES 4 4 Phenylalanine amide.
 SQ SEQUENCE 4 AA; 598 MW; 69D4073B30000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 4;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 X 1
 DB 1 Y 1

RESULT 13
 PAR4_HIRME STANDARD; PRT; 4 AA.
 AC P42563;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE FMRamide-like neuropeptide YLRF-amide.
 OS Hirudo medicinalis (Medicinal leech).
 OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinidae; Hirudo.
 OX NCBI_TaxID=6421;
 RN [1]
 RP SEQUENCE.
 RA MEDLINE=92195954; PubMed=1686933; DOI=10.1016/0196-9781(91)90035-N;
 RX Evans B.D., Pohl J., Kartsonis M.A., Calabrese R.L.;
 RT "Identification of Rfamide neuropeptides in the medicinal leech.";
 RL Peptides 12:897-908(1991).
 CC -|- SUBCELLULAR LOCATION: Secreted.
 CC -|- SIMILARITY: Belongs to the FARP (FMRamide related peptide) family.
 KW Amidation; Direct protein sequencing; Neuropeptide.
 FT MOD RES 4 4 Phenylalanine amide.
 SQ SEQUENCE 4 AA; 616 MW; 69D4068B30000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 4;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 X 1
 DB 1 Y 1

RESULT 14
 PFKA_ANTEL STANDARD; PRT; 4 AA.
 AC P58705;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Antho-Kamide.
 OS Anthopleura elegantissima (Sea anemone).
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria; Nynanthaeae; Actinidae; Anthopleura.

OX NCBI_TaxID=6110;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=92028852; PubMed=1681803;
 RA Nothacker H.-P., Rinehart K.L. Jr., Grimmelikhuijzen C.J.P.;
 RT "Isolation of L-3-phenyllactyl-Phe-Lys-Ala-NH2 (Antho-Kamide), a novel neuropeptide from sea anemones";
 RL Biochem. Biophys. Res. Commun. 179:1205-1211(1991).
 RN [2]
 RP FUNCTION.
 RA MEDLINE=93391436; PubMed=8397415;
 RX McFarlane I.D., Hudman D., Nothacker H.-P., Grimmelikhuijzen C.J.P.;
 RT "The expansion behaviour of sea anemones may be coordinated by two inhibitory neuropeptides, Antho-Kamide and Antho-Ramide.";
 RL Proc. R. Soc. Lond., B, Biol. Sci. 253:183-188(1993).
 CC -|- FUNCTION: Inhibits spontaneous contractions in several muscle groups. May be involved in the expansion phase of feeding behaviour in sea anemones.
 CC -|- SUBCELLULAR LOCATION: Secreted.
 CC -|- TISSUE SPECIFICITY: Neuron specific.
 DR PIR; JQ1273; JQ1273.
 KW Amidation; Direct protein sequencing; Neuropeptide.
 FT MOD RES 1 1 3-phenyllactic acid.
 FT MOD RES 4 4 Alanine amide.
 SQ SEQUENCE 4 AA; 512 MW; 6DD339C9A0000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 4;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 X 1
 DB 4 A 4

RESULT 15
 FLRF_HIRME STANDARD; PRT; 4 AA.
 AC P42561;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE FLRFamide.
 OS Hirudo medicinalis (Medicinal leech), and
 OS Helisoma trivolvis (Snail).
 OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinidae; Hirudo.
 OX Arhynchobdellida; Hirudiniformes; Hirudinidae; Hirudo.
 OX NCBI_TaxID=6421, 27815;
 RN [1]
 RP SEQUENCE.
 RC SPECIES=H.medicalinalis;
 RX MEDLINE=92195954; PubMed=1686933; DOI=10.1016/0196-9781(91)90035-N;
 RA Evans B.D., Pohl J., Kartsonis M.A., Calabrese R.L.;
 RT "Identification of Rfamide neuropeptides in the medicinal leech.";
 RL Peptides 12:897-908(1991).
 RN [2]
 RP SEQUENCE.
 RC SPECIES=H.trivolvis; TISSUE=Kidney;
 RX MEDLINE=94286417; PubMed=7912428; DOI=10.1016/0196-9781(94)90166-X;
 RA Madrid K.P., Price D.A., Greenberg M.J., Khan H.R., Saleuddin A.S.M.;
 RT "FMRamide-related peptides from the kidney of the snail, Helisoma trivolvis.";
 RL Peptides 15:31-36(1994).
 CC -|- SUBCELLULAR LOCATION: Secreted.
 CC -|- SIMILARITY: Belongs to the FARP (FMRamide related peptide) family.
 KW Amidation; Direct protein sequencing; Neuropeptide.
 FT MOD RES 4 4 Phenylalanine amide.
 SQ SEQUENCE 4 AA; 582 MW; 69D40729A0000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 4;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 1 F 1

RESULT 16

FLRN ANTEL
ID FLRN ANTEL STANDARD; PRT; 4 AA.
AC P58707;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Antho-RNamide.
OS Anthopleura elegantissima (Sea anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nynanthaea; Actiniidae; Anthopleura.
OX NCBI_TaxID=6110;
RN [1]
RN SEQUENCE, AND MASS SPECTROMETRY.
RX MEDLINE=90319122; PubMed=1973541;
RA Grimmelikhuijzen C.J.P., Rinehart K.L. Jr., Jacob E., Graff D.,
RA Reinscheid R.K., Nothacker H.-P., Staley A.L.;
RT "Isolation of L-3-phenylacetyl-Leu-Arg-Asn-NH2 (Antho-RNamide), a sea
anemone neuropeptide containing an unusual amino-terminal blocking
group.";
RT Proc. Natl. Acad. Sci. U.S.A. 87:5410-5414 (1990).
RL
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- TISSUE SPECIFICITY: Neuron specific.
CC -|- MASS SPECTROMETRY: MW=549.3; METHOD=PAB; RANGE=1-4; NOTE=Ref.1.
DR PIR; A35779; A35779.
KW Amidation; Direct protein sequencing; Neuropeptide.
FT MOD_RES 1 1 3-phenylacetic acid.
FT MOD_RES 4 4 Asparagine amide.
SQ SEQUENCE 4 AA; 549 MW; 64540729A0000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 1 F 1

RESULT 17

FMRF MACNI
ID FMRF MACNI STANDARD; PRT; 4 AA.
AC P01162;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE FMRFamide (Peak C) (Cardioexcitatory neuropeptide).
OS Macrocyclista nimbosa (Sun-ray clam),
OS Neris virgens (Sandworm),
OS Hirudo medicinalis (Medicinal leech), and
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Heteroconchia; Veneroidea;
OC Veneroidea; Veneridae; Macrocallista.
OX NCBI_TaxID=6594, 6353, 6421, 27815;
RN [1]
RN SEQUENCE, AND SYNTHESIS.
RC SPECIES=M.nimbosa; TISSUE=Cerebral pedal, and Visceral ganglion;
RX MEDLINE=77215956; PubMed=877582;
RA Price D.A., Greenberg M.J.;
RT "Structure of a molluscan cardioexcitatory neuropeptide.";
RL Science 197:670-671 (1977).
RN [2]

RP SEQUENCE, AND CHARACTERIZATION.
RC SPECIES=M.nimbosa; TISSUE=Ganglion;
RX MEDLINE=78012038; PubMed=909875;
RA Price D.A., Greenberg M.J.;
RT "Purification and characterization of a cardioexcitatory neuropeptide

RT from the central ganglia of a bivalve mollusc.";
RN Prep. Biochem. 7:261-281 (1977).
RP SEQUENCE.
RC SPECIES=N.virens;
RX MEDLINE=90259866; PubMed=2342992; DOI=10.1016/0196-9781(90)90113-J;
RA Krajciak K.G., Price D.A.;
RT "Authentic FMRFamide is present in the polychaete Nereis virens.";
RL Peptides 11:75-77 (1990).
RN [4]
RN SEQUENCE.
RP SEQUENCE.
RX MEDLINE=92195954; PubMed=1686933; DOI=10.1016/0196-9781(91)90035-N;
RA Evans B.D., Pohl J., Kartsonis M.A., Calabrese R.L.;
RT "Identification of RFamide neuropeptides in the medicinal leech.";
RL Peptides 12:897-908 (1991).
RN [5]
RN SEQUENCE.
RP SEQUENCE.
RX MEDLINE=H.trivolvus; TISSUE=Kidney;
RX MEDLINE=94286417; PubMed=7912428; DOI=10.1016/0196-9781(94)90166-X;
RA Madrid K.P., Price D.A., Greenberg M.J., Khan H.R., Saleuddin A.S.M.;
RT "FMRFamide-related peptides from the kidney of the snail, Helisoma
trivolvus.";
RL Peptides 15:31-36 (1994).
CC -|- FUNCTION: Myoactive; cardioexcitatory substance. Pharmacological
activities include augmentation, induction, and regularization of
cardiac contraction.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- SIMILARITY: Belongs to the FARP (FMRFamide related peptide)
family.
DR PIR; A01426; ECNK.
DR PIR; A60418; A60418.
KW Amidation; Direct protein sequencing; Neuropeptide.
FT MOD_RES 4 4 Phenylalanine amide.
SQ SEQUENCE 4 AA; 600 MW; 69D40699A0000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 1 F 1

RESULT 18

FYRI ANTEL
ID FYRI ANTEL STANDARD; PRT; 4 AA.
AC P58706;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Antho-RNamide I [Contains: Antho-RNamide II].
OS Anthopleura elegantissima (Sea anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nynanthaea; Actiniidae; Anthopleura.
OX NCBI_TaxID=6110;
RN [1]
RN SEQUENCE.
RX MEDLINE=92270459; PubMed=1821096; DOI=10.1016/0196-9781(91)90190-Z;
RA Nothacker H.-P., Rinehart K.L. Jr., McFarlane I.D.,
RA Grimmelikhuijzen C.J.P.;
RT "Isolation of two novel
biologically active L-3-phenylacetyl-Tyr-Arg-Ile-NH2 and its des-
phenylacetyl fragment Tyr-Arg-Ile-NH2.";
RL Peptides 12:1165-1173 (1991).
RN [2]
RN FUNCTION.
RX MEDLINE=93391436; PubMed=8397415;
RA McFarlane I.D., Hudman D., Nothacker H.-P., Grimmelikhuijzen C.J.P.;
RT "The expansion behaviour of sea anemones may be coordinated by two
inhibitory neuropeptides, Antho-RNamide and Antho-RNamide.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 253:183-188 (1993).

CC -!- FUNCTION: Inhibits spontaneous contractions in several muscle
CC groups. May be involved in the expansion phase of feeding
CC behaviour in sea anemones.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Neuron specific.
KW Amidation; Direct protein sequencing; Neuropeptide.
FT CHAIN 1 4 Antho-Ramide I.
FT CHAIN 2 4 Antho-Ramide II.
FT MOD RES 1 1 3-phenyllactic acid.
FT MOD RES 4 4 Isoleucine amide.
SQ SEQUENCE 4 AA; 598 MW; 60441B59A0000000 CRC64;
Query Match 0.0%; Score 0; DB 1; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 X 1
Db 1 F 1
RESULT 19
ILME SEPOF
ID ILME SEPOF STANDARD; PRT; 4 AA.
AC P83568;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Phormone peptide ILME.
OS Sepia officinalis (Common cuttlefish).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
OC Decapodiformes; Sepioidea; Sepiidae; Sepia.
OX NCBI_TaxID=6610;
RN [1]
RP SEQUENCE, SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND MASS
RP SPECTROMETRY.
RC TISSUE=Egg;
RX MEDLINE=20403899; PubMed=10944467; DOI=10.1006/bbrc.2000.3286;
RA Zatylny C., Gagnon J., Boucaud-Camou E., Henry J.;
RT "ILME: a waterborne pheromonal peptide released by the eggs of Sepia
RT officinalis.";
RL Biochem. Biophys. Res. Commun. 275:217-222(2000).
RN [2]
RP SEQUENCE.
RC TISSUE=Egg;
RX MEDLINE=22197108; PubMed=12207899; DOI=10.1016/S0006-291X(02)02036-3;
RA Zatylny C., Marvin L., Gagnon J., Henry J.;
RT "Fertilization in Sepia officinalis: the first mollusk sperm-
RT attracting peptide.";
RL Biochem. Biophys. Res. Commun. 296:1186-1193(2002).
CC -!- FUNCTION: Has myotropic activity targeting the genital tract.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Follicle, fully grown oocyte and egg (EC2).
CC -!- MASS SPECTROMETRY: MW=505.4; METHOD=MALDI; RANGE=1-4; NOTE=Ref.1.
KW Direct protein sequencing; Pheromone.
SQ SEQUENCE 4 AA; 505 MW; 6B16972030000000 CRC64;
Query Match 0.0%; Score 0; DB 1; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 X 1
Db 1 I 1
RESULT 20
OCP1_OCTMI
ID OCP1_OCTMI STANDARD; PRT; 4 AA.
AC P58648;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Cardioactive peptides Ocp-1/Ocp-2.
OS Octopus minor (Octopus).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
OC Octopodiformes; Octopoda; Incirrata; Octopodidae; Octopus.
OX NCBI_TaxID=89766;
RN [1]
RP SEQUENCE, SYNTHESIS, MASS SPECTROMETRY, AND CHARACTERIZATION.
RC TISSUE=Brain;
RC MEDLINE=20336815; PubMed=10876044; DOI=10.1016/S0196-9781(00)00201-1;
RA Iwakoshi E., Hisada M., Minakata H.;
RT "Cardioactive peptides isolated from the brain of a Japanese octopus,
RT Octopus minor.";
RL Peptides 21:623-630(2000).
CC -!- FUNCTION: Cardioactive; has both positive chronotropic and
CC inotropic effects on the heart. Ocp-2 is a 1000 time less active
CC than Ocp-1.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- PTM: Ocp-2 has L-Phe instead of D-Phe.
CC -!- MASS SPECTROMETRY: MW=395.2; METHOD=MALDI; RANGE=1-4; NOTE=Ref.1.
KW D-amino acid; Direct protein sequencing; Hormone.
FT MOD RES 2 2 D-phenylalanine (in form Ocp-1).
SQ SEQUENCE 4 AA; 394 MW; 6AA879C810000000 CRC64;
Query Match 0.0%; Score 0; DB 1; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 X 1
Db 1 G 1
RESULT 21
OCP3_OCTMI
ID OCP3_OCTMI STANDARD; PRT; 4 AA.
AC P58649;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Cardioactive peptides Ocp-3/Ocp-4.
OS Octopus minor (Octopus).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
OC Octopodiformes; Octopoda; Incirrata; Octopodidae; Octopus.
OX NCBI_TaxID=89766;
RN [1]
RP SEQUENCE, SYNTHESIS, MASS SPECTROMETRY, AND CHARACTERIZATION.
RC TISSUE=Brain;
RC MEDLINE=20336815; PubMed=10876044; DOI=10.1016/S0196-9781(00)00201-1;
RA Iwakoshi E., Hisada M., Minakata H.;
RT "Cardioactive peptides isolated from the brain of a Japanese octopus,
RT Octopus minor.";
RL Peptides 21:623-630(2000).
CC -!- FUNCTION: Cardioactive; has both positive chronotropic and
CC inotropic effects on the heart. Ocp-4 is a 1000 time less active
CC than Ocp-3.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- PTM: Ocp-4 has D-Ser instead of L-Ser.
CC -!- MASS SPECTROMETRY: MW=395.2; METHOD=MALDI; RANGE=1-4; NOTE=Ref.1.
KW D-amino acid; Direct protein sequencing; Hormone.
FT MOD RES 2 2 D-serine (in form Ocp-4).
SQ SEQUENCE 4 AA; 463 MW; 6AB365B810000000 CRC64;
Query Match 0.0%; Score 0; DB 1; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 X 1
Db 2 S 2
RESULT 22
TUFT_HUMAN

ID TUFT HUMAN STANDARD; PRT; 4 AA.
AC P01858;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Phagocytosis-stimulating peptide (tuftsin).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=72187087; PubMed=4112769;
RA Nishioka K., Constantopoulos A., Satoh P.S., Najjar V.A.;
RT "The characteristics, isolation and synthesis of the phagocytosis
stimulating peptide tuftsin.";
RL Biochem. Biophys. Res. Commun. 47:172-179(1972).
RN [2]
RP IMMUNOGLOBULIN CLASS.
RX MEDLINE=68091045; PubMed=4169272;
RA Fidalgo B.V., Najjar V.A.;
RT "The physiological role of the lymphoid system. VI. The stimulatory
effect of leucophilic gamma globulin (leucokinin) on the phagocytic
activity of human polymorphonuclear leucocyte.";
RL Biochemistry 6:3386-3392(1967).
CC -!- MISCELLANEOUS: An IgG (called leucokinin) binds reversibly to the
cell membrane of neutrophils in the blood. Leucokininase on the
membrane releases the active peptide tuftsin from the gamma chain.
CC Tuftsin is essential for maximum stimulation of the phagocytic
activity of neutrophils.
CC PIR; A02147; A02147.
DR MIM; 191150;
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006909; P:phagocytosis; NAS.
KW Direct protein sequencing.
SQ SEQUENCE 4 AA; 501 MW; 74176331C00000000 CRC64;
Query Match 0.0%; Score 0; DB 1; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 X 1
Db 1 T 1
RESULT 23
VLM1 YEAST
ID YLM1 YEAST STANDARD; PRT; 4 AA.
AC P36515;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Mitochondrial 60S ribosomal protein L1 (Yml1) (Fragment).
GN Name=Yml1;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE.
RX MEDLINE=91285106; PubMed=2060626; DOI=10.1016/0014-5793(91)80759-V;
RA Grohmann L., Graack H.-R., Kruft V., Choli T., Goldschmidt-Reisin S.,
RA Kitakawa M.;
RT "Extended N-terminal sequencing of proteins of the large ribosomal
subunit from yeast mitochondria.";
RL FEBS Lett. 284:51-56(1991).
CC -!- FUNCTION: Putative component of the large subunit of mitochondrial
ribosome.
CC -!- SUBCELLULAR LOCATION: Mitochondrial.
CC PIR; S17255; S17255.
KW Direct protein sequencing; Mitochondrion; Ribosomal protein.
FT : NON_TER 4

SQ SEQUENCE 4 AA; 402 MW; 7771B2D5D00000000 CRC64;
Query Match 0.0%; Score 0; DB 1; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 X 1
Db 1 S 1
RESULT 24
Q16047
ID Q16047 PRELIMINARY; PRT; 4 AA.
AC Q16047;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Interleukin 2 receptor alpha-subunit (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92062171; PubMed=1840490;
RA Mercken L., Moras V., Hemon L., Lionne B., Bousseau A.,
RA Dautry-Varsat A., Collins M., Mayaux J.F.;
RT "An exon 5-deleted mRNA encodes a functional interleukin 2 receptor
alpha-subunit.";
RL Biochem. Biophys. Res. Commun. 180:1390-1395(1991).
DR EMBL; S64248; AAB20279.1;
DR HSSP; P01112; 1PLL.
DR GO; GO:0005525; F:GTP binding; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR InterPro; IPR001806; Ras trnsfrmg.
DR PRINTS; PR00449; RASTRNSFRMG.
KW Receptor.
FT NON_TER 4
SQ SEQUENCE 4 AA; 525 MW; 69CAB769A00000000 CRC64;
Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 X 1
Db 1 F 1
RESULT 25
Q96AT0
ID Q96AT0 PRELIMINARY; PRT; 4 AA.
AC Q96AT0;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE FLJ30856 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyzinski M.I., Skalska U., Smalilov D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC016791; AAH16791.2; -;
SQ SEQUENCE 4 AA; 512 MW; 633DCB56F00000000 CRC64;

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 3 A 3

RESULT 26
Q08433
ID Q08433 PRELIMINARY; PRT; 4 AA.
AC Q08433;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Bilirubin UDP-glucuronosyltransferase (Fragment).
OS Rattus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10118;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Gunn;
RA MEDLINE=91282758; PubMed=1840486;
RA Sato H., Aono S., Kashiwamata S., Koiwai O.;
RT "Genetic defect of bilirubin UDP-glucuronosyltransferase in the
RT hyperbilirubinemic Gunn rat.";
RL Biochem. Biophys. Res. Commun. 177:1161-1164(1991).
DR EMBL, S38636; AAB19259.1; -;
DR GO; GO:0016740; F:transferase activity; IEA.
KW Transferase.
FT NON_TER 4 4
SQ SEQUENCE 4 AA; 473 MW; 633732C4200000000 CRC64;

Query Match 0.0%; Score 0; DB 2; Length 4;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 1 N 1

RESULT 27
AL14_CARMA
ID AL14_CARMA STANDARD; PRT; 5 AA.
AC P81817;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Adenosylmethionine-8-amino-7-oxononanoate aminotransferase
DE (EC 2.6.1.62) (7,8-diamino-pelargonic acid aminotransferase) (DAPA

DE Carcinustatin 14.
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;
RX MEDLINE=98121193; PubMed=9461295;
RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA Thorpe A.;
RT "Isolation and identification of multiple neuropeptides of the
RT allatostatin superfamily in the shore crab Carcinus maenas.";
RL Eur. J. Biochem. 250:727-734(1997).
CC -!- FUNCTION: May act as a neurotransmitter or neuromodulator.
CC -!- SIMILARITY: Belongs to the allatostatin family.
KW Amidation; Direct protein sequencing; Multigene family; Neuropeptide.
FT MOD_RES 5 5
SQ SEQUENCE 5 AA; 586 MW; 672879D5AB3000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 5;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 2 S 2

RESULT 28
AP21_EISFO
ID AP21_EISFO STANDARD; PRT; 5 AA.
AC P84182;
DT 25-JAN-2005 (Rel. 46, Created)
DT 25-JAN-2005 (Rel. 46, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Antimicrobial peptide OEP3121.
OS Eisenia foetida (Common dung-bird worm) (Common dung-worm).
OC Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxida;
OC Lumbricina; Lumbricidae; Eisenia.
OX NCBI_TaxID=6396;
RN [1]
RP SEQUENCE, FUNCTION, AND MASS SPECTROMETRY.
RX PubMed=15253156;
RA Liu Y.-Q., Sun Z.-J., Wang C., Li S.-J., Liu Y.-Z.;
RT "Purification of a novel antibacterial short peptide in earthworm
RT Eisenia foetida.";
RL Acta Biochim. Biophys. Sin. 36:297-302(2004).
CC -!- FUNCTION: Displays antimicrobial activity against E.coli, S.aureus
CC and P.aeruginosa.
CC -!- MASS SPECTROMETRY: MW=510.8; METHOD=NALDI; RANGE=1-5; NOTE=Ref.1.
KW Antibiotic; Direct protein sequencing.
SQ SEQUENCE 5 AA; 407 MW; 687DC5BEBDB000000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 5;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 1 A 1

RESULT 29
BIOA_CITFR
ID BIOA_CITFR STANDARD; PRT; 5 AA.
AC P13071;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Adenosylmethionine-8-amino-7-oxononanoate aminotransferase
DE (EC 2.6.1.62) (7,8-diamino-pelargonic acid aminotransferase) (DAPA

DE aminotransferase) (Fragment).
GN Name=bioA;
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89006280; PubMed=2971595; DOI=10.1016/0378-1119(88)90397-6;
RA Shiuan D., Campbell A.;
RT "Transcriptional regulation and gene arrangement of Escherichia coli,
Citrabacter freundii and Salmonella typhimurium biotin operons.";
RL Gene 67:203-211(1988).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + 8-amino-7-oxononanoate = S-adenosyl-4-methylthio-2-oxobutanoate + 7,8-diaminononanoate.
CC -!- COFACTOR: Pyridoxal phosphate.
CC -!- PATHWAY: Biotin biosynthesis.
CC -!- SUBUNIT: Homodimer.
CC -!- SIMILARITY: Belongs to the class-III pyridoxal-phosphate-dependent aminotransferase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M21922; -; NOT_ANNOTATED_CDS.
DR PIR; I40697; I40697.
DR InterPro; IPR005814; AminoTrans 3.
DR PROSITE; PS00600; AA_TRANSFER_CLASS 3; PARTIAL.
KW Aminotransferase; Biotin biosynthesis; Pyridoxal phosphate;
KW Transfrase.
FT NON_TER 5
SQ SEQUENCE 5 AA; 582 MW; 6AAAB1B1A6F00000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 5;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 2 T 2

RESULT 30
BIOB_CITFR
ID BIOB_CITFR STANDARD; PRT; 5 AA.
AC P12997;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Biotin synthase (EC 2.8.1.6) (Biotin synthetase) (Fragment).
GN Name=bioB;
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89006280; PubMed=2971595; DOI=10.1016/0378-1119(88)90397-6;
RA Shiuan D., Campbell A.;
RT "Transcriptional regulation and gene arrangement of Escherichia coli,
Citrabacter freundii and Salmonella typhimurium biotin operons.";
RL Gene 67:203-211(1988).
CC -!- CATALYTIC ACTIVITY: Dethiobiotin + sulfur = biotin.
CC -!- COFACTOR: Binds a 4Fe-4S cluster coordinated with 3 cysteines and an exchangeable S-adenosyl-L-methionine, and a 2Fe-2S cluster coordinated with 3 cysteines and 1 arginine (By similarity).
CC -!- PATHWAY: Biotin biosynthesis; last step.

CC -!- SIMILARITY: Belongs to the biotin and lipoic acid synthetases family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M21922; -; NOT_ANNOTATED_CDS.
DR PIR; I40698; I40698.
DR 2Fe-2S; 4Fe-4S; Biotin biosynthesis; Iron-sulfur; Transfrase.
FT NON_TER 5
SQ SEQUENCE 5 AA; 532 MW; 75A5B1EDD6F00000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 5;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 2 A 2

RESULT 31
BPP7_BOTIN
ID BPP7_BOTIN STANDARD; PRT; 5 AA.
AC P30425;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Bradykinin-potentiating peptide S5,2 (5A) (Angiotensin-converting enzyme inhibitor).
DE Enzyme inhibitor).
OS Bothrops insularis (Island jararaca) (Queimada jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Bothrops.
OX NCBI_TaxID=8723;
RN [1]
RP SEQUENCE.
RC TISSUE=Venom;
RX MEDLINE=90351557; PubMed=2386615;
RA Cintra A.C.O., Vieira C.A., Giglio J.R.;
RT "Primary structure and biological activity of bradykinin potentiating peptides from Bothrops insularis snake venom."
RL J. Protein Chem. 9:221-227(1990).
CC -!- FUNCTION: This peptide both inhibits the activity of the angiotensin-converting enzyme and enhances the action of bradykinin by inhibiting the kinases that inactivate it. It acts as an indirect hypotensive agent.
CC PIR; G37196; G37196.
DR Direct protein sequencing; Hypotensive agent;
KW Pyrrolidone carboxylic acid.
FT MOD_RES 1
SQ SEQUENCE 5 AA; 629 MW; 776DC37326B00000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 5;
Best Local Similarity 0.0%; Pred.No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
Db 4 A 4

RESULT 32
EI03_LITRU
ID EI03_LITRU STANDARD; PRT; 5 AA.
AC P82099;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Electrin 3.
OS Litoria rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Hyloidea; Hylidae;
OC Pelodyadinae; Litoria.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE.
RC TISSUE=Skin secretion;
RA Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;
RT "Peptides from the skin glands of the Australian buzzing tree frog
RT Litori electrica. Comparison with the skin peptides from Litoria
RT rubella".
RL Aust. J. Chem. 52:639-645(1999).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Skin.
KW Amidation; Amphibian defense peptide; Direct protein sequencing.
FT MOD_RES 5 5 Methionine amide.
SQ SEQUENCE 5 AA; 630 MW; 668761F2C9A00000 CRC64;
Query Match 0.0%; Score 0; DB 1; Length 5;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 X 1
DB 1 F 1
RESULT 33
ID E104_LITRU STANDARD; PRT; 5 AA.
AC P82100;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Electrin 4.
OS Litoria rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Hyloidea; Hylidae;
OC Pelodyadinae; Litoria.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE.
RC TISSUE=Skin secretion;
RA Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;
RT "Peptides from the skin glands of the Australian buzzing tree frog
RT Litori electrica. Comparison with the skin peptides from Litoria
RT rubella".
RL Aust. J. Chem. 52:639-645(1999).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Skin.
KW Amidation; Amphibian defense peptide; Direct protein sequencing.
FT MOD_RES 5 5 Histidine amide.
SQ SEQUENCE 5 AA; 616 MW; 61F2D1A059A00000 CRC64;
Query Match 0.0%; Score 0; DB 1; Length 5;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 X 1
DB 3 T 3
RESULT 34
ID FARP_ARTTR STANDARD; PRT; 5 AA.
AC P41853;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE FMRFamide-like neuropeptide RVIRF-amide.
OS Artioptothia triangulata (New Zealand flatworm).
OC Eukaryota; Metazoa; Platyhelminthes; Turbellaria; Seriata; Tricladida;
OC Terricola; Geoplanidae; Arthurdendyus.
OX NCBI_TaxID=132421;
RN [1]
RP SEQUENCE, AND SYNTHESIS
RX MEDLINE=94211927; PubMed=7909164; DOI=10.1016/0167-0115(94)90189-9;
RA Maule A.G., Shaw C., Halton D.W., Curry W.J., Thim L.;
RT "RVIRFamide: a turbellarian FMRFamide-related peptide (FARP).";
RL Regul. Pept. 50:37-43(1994).
CC -!- SIMILARITY: Belongs to the FARP (FMRFamide related peptide)
CC family.
KW Amidation; Direct protein sequencing; Neuropeptide.
FT MOD_RES 5 5 Phenylalanine amide.
SQ SEQUENCE 5 AA; 754 MW; 69D4004B44600000 CRC64;
Query Match 0.0%; Score 0; DB 1; Length 5;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 X 1
DB 1 R 1
RESULT 35
ID FARP_CHICK STANDARD; PRT; 5 AA.
AC P83308;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE FMRFamide-like neuropeptide (LPLRF-amide).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE, AND SYNTHESIS.
RC TISSUE=Brain;
RX PubMed=6137771;
RA Dockray G.J., Reeve J.R. Jr., Shively J., Gayton R.J., Barnard C.S.;
RT "A novel active pentapeptide from chicken brain identified by
RT antibodies to FMRFamide".
RL Nature 305:328-330(1983).
CC -!- FUNCTION: May function as a neurotransmitter or modulator.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the FARP (FMRFamide related peptide)
CC family.
DR GO:0007218; P:neuropeptide signaling pathway; TAS.
KW Amidation; Direct protein sequencing; Neuropeptide.
FT MOD_RES 5 5 Phenylalanine amide.
SQ SEQUENCE 5 AA; 645 MW; 69D4073767400000 CRC64;
Query Match 0.0%; Score 0; DB 1; Length 5;
Best Local Similarity 0.0%; Pred. No. 0;
Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 X 1
DB 1 L 1
RESULT 36
ID MPA4_JUNVI STANDARD; PRT; 5 AA.
AC P81826;
DT 25-OCT-2004 (Rel. 45, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Major pollen allergen Jun v 4 (Fragment).

OS Juniperus virginiana (Eastern red cedar).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Coniferopsida; Coniferales; Cupressaceae; Juniperus.
 OX NCBI_TaxID=39584;
 RN [1]
 RP SEQUENCE, AND ALLERGENIC PROPERTIES.
 RC TISSUE=Pollen;
 RX MEDLINE=21315424; PubMed=11422137;
 RA Midoro-Horiuti T., Goldblum R.M., Brooks E.G.;
 RT "Identification of mutations in the genes for the pollen allergens of
 RT eastern red cedar (Juniperus virginiana).";
 RL Clin. Exp. Allergy 31:771-778(2001).
 CC -!- ALLERGEN: Causes an allergic reaction in human. Binds to IgE of
 CC patients who are allergic to J.ashei.
 KW Allergen; Direct protein sequencing.
 FT NON_TER 5
 SQ SEQUENCE 5 AA; 491 MW; 75B33DDAADB00000 CRC64;
 QY 1 X 1
 Db 1 A 1
 RESULT 37
 PAPT2_PAPMA STANDARD; PRT; 5 AA.
 ID P81864;
 AC 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Pardaxin II (PXII) (Fragment).
 OS Pardachirus marmoratus (Red sea mores sole).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
 OC Soleiidae; Soleidae; Pardachirus.
 OX NCBI_TaxID=31087;
 RN [1]
 RP TISSUE=Skin secretion;
 RC MEDLINE=87057369; PubMed=3782138;
 RA Lazarovici P., Primor N., Loew L.M.;
 RT "Purification and pore-forming activity of two hydrophobic
 RT polypeptides from the secretion of the Red sea mores sole (Pardachirus
 RT marmoratus).";
 RL J. Biol. Chem. 261:16704-16713(1986).
 CC -!- FUNCTION: Exhibits unusual shark repellent and surfactant
 CC properties. Forms voltage-dependent, ion-permeable channels in
 CC membranes. At high concentration causes cell membrane lysis.
 CC -!- SUBUNIT: Monomer. In aqueous solution exists as a tetramer.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the pardaxin family.
 KW Direct protein sequencing; Toxin.
 FT NON_TER 5
 SQ SEQUENCE 5 AA; 614 MW; 7769C9C8100000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 5;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
 Db 1 G 1

RESULT 38
 PAPT2_PAPMA STANDARD; PRT; 5 AA.
 ID P67857; P01373;

DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Proctolin.
 OS Carcinus maenas (Common shore crab) (Green crab).
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
 OC Eubrachyura; Portunoidae; Portunidae; Carcinus.
 OX NCBI_TaxID=6759;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=8623789; PubMed=2872661; DOI=10.1016/0196-9781(86)90063-X;
 RA Stangier J., Dirksen H., Keller R.;
 RT "Identification and immunocytochemical localization of proctolin in
 RT pericardial organs of the shore crab, Carcinus maenas.";
 RL Peptides 7:67-72(1986).
 CC -!- FUNCTION: Stimulates cardiac output and hindgut motility,
 CC modulates visceral and skeletal muscle in many arthropods.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Found in the crab pericardial organs.
 KW Direct protein sequencing; Neuropeptide.
 SQ SEQUENCE 5 AA; 649 MW; 71B7673B44600000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 5;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
 Db 5 T 5

RESULT 39
 PAPT2_LIMPO STANDARD; PRT; 5 AA.
 ID P67858; P01373;
 AC 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Proctolin.
 OS Limulus polyphemus (Atlantic horseshoe crab).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
 OC Limulidae; Limulus.
 OX NCBI_TaxID=6850;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=90287800; PubMed=2356151; DOI=10.1016/0196-9781(90)90072-D;
 RA Groomer J.R., Tillinghast E.K., Townley M.A., Vetrovs A.,
 RA Watson W.H. III, Hunt D.F., Griffin P.R., Alexander J.E.,
 RA Shabanowitz J.;
 RT "Identification of proctolin in the central nervous system of the
 RT horseshoe crab, Limulus polyphemus.";
 RL Peptides 11:205-211(1990).
 CC -!- FUNCTION: Stimulates cardiac output and hindgut motility,
 CC modulates visceral and skeletal muscle in many arthropods.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Found in the crab pericardial organs.
 KW PIR: A60411; A60411.
 KW Direct protein sequencing; Neuropeptide.
 SQ SEQUENCE 5 AA; 649 MW; 71B7673B44600000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 5;
 Best Local Similarity 0.0%; Pred. No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 X 1
 Db 5 T 5

RESULT 40
 PAPT2_PERAM STANDARD; PRT; 5 AA.
 ID P67857; P01373;

AC P67859; P01373;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Proctolin.
 OS Periplaneta americana (American cockroach).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Orthopteroidea; Dictyoptera; Blattaria; Blattodea;
 OC Blattidae; Periplaneta.
 OX NCBI_TaxID=6978;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=76074708; PubMed=576; DOI=10.1016/0024-3205(75)90134-4;
 RA Starratt A.N., Brown B.E.;
 RT "Structure of the pentapeptide proctolin, a proposed neurotransmitter
 in insects.";
 RL Life Sci. 17:1253-1256(1975).
 RN [2]
 RP BIOLOGICAL SOURCE.
 RX MEDLINE=81225865; PubMed=6113690;
 RA O'Shea M., Adams M.E.;
 RT "Pentapeptide (proctolin) associated with an identified neuron.";
 RL Science 213:567-569(1981).
 CC -!- FUNCTION: Stimulates cardiac output and hindgut motility,
 CC modulates visceral and skeletal muscle in many arthropods.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Found in the lateral white neurons.
 DR PIR; A01644; HOROHA.
 KW Direct protein sequencing; Neuropeptide.
 SQ SEQUENCE 5 AA; 649 MW; 71B7673B44600000 CRC64;

Query Match 0.0%; Score 0; DB 1; Length 5;
 Best Local Similarity 0.0%; Pred.No. 0;
 Matches 0; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 X 1
 Db 5 T 5

Search completed: April 13, 2005, 17:16:22
 Job time : 13.5912 secs

This Page Blank (uspto)